

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:16:43 ; Search time 873 seconds
(without alignments)
8912.548 Million cell updates/sec

Title: US-10-072-621-4.rng
Perfect score: 3455
Sequence: 1 gccaccatgcccgaagc.....atgccagtgtattgtggg 3455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

ched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2207.8	63.9	2258	21 AA294756	Human ATP binding
2	1931.4	55.9	1941	24 AAL44669	Human transporter
3	1062.4	30.7	3376	22 AAK51722	Human polynucleoti
4	1062.4	30.7	3376	22 AAH90113	Human bone marrow
5	1042.2	30.2	3586	24 AB199363	Mouse ischaemic co
6	972.8	28.2	2930	21 AA294747	Human ATP binding
7	972.8	28.2	2930	24 ABL63321	Breast cancer rela
8	533.2	15.4	687	24 AB072606	Human MDPT encodin
9	396.8	11.5	2208	23 ABL10101	Drosophila melanog

10	391.8	11.3	2283	23 ABL05663	Drosophila melanog
11	332.8	9.6	2639	23 ABL15499	Drosophila melanog
12	324.2	9.4	2537	21 AA298624	Silkworm Bm white
13	313.6	9.1	4497	23 ABL05455	Drosophila melanog
14	253.6	7.3	2182	23 ABL15815	Drosophila melanog
15	243.6	7.1	13815	23 ABL05454	Drosophila melanog
16	238	6.9	287	22 AA20365	Human brain expres
17	238	6.9	287	22 AAK46460	Human bone marrow
18	238	6.9	287	24 ABS20791	Human genome-deriv
19	220.8	6.4	2396	23 ABL08195	Drosophila melanog
20	220.8	6.4	2590	23 ABL20321	Drosophila melanog
21	220.8	6.4	2894	23 ABL20331	Drosophila melanog
22	220	6.4	2352	23 ABL05135	Drosophila melanog
23	217	6.3	491	22 AAK07701	Human brain expres
24	217	6.3	491	22 AAK33523	Human bone marrow
25	217	6.3	491	24 ABS08376	Human genome-deriv
26	207.6	6.0	2576	23 ABL15497	Drosophila melanog
27	203.2	5.9	5310	23 ABL15498	Drosophila melanog
28	198	5.7	2367	23 ABL22265	Drosophila melanog
29	169.8	4.9	2525	21 AA298625	Silkworm Bm white
30	163.4	4.7	2325	23 ABL20329	Drosophila melanog
31	160.6	4.6	8361	23 ABL05662	Drosophila melanog
32	154.6	4.5	1794	23 ABL12777	Drosophila melanog
33	135.6	3.9	5685	23 ABL12776	Drosophila melanog
34	129	3.7	8415	23 ABL15496	Drosophila melanog
35	110.8	3.2	10386	23 ABL22264	Drosophila melanog
36	107.8	3.1	5383	23 ABL20328	Drosophila melanog
37	107.8	3.1	5503	23 ABL20320	Drosophila melanog
38	107.8	3.1	5506	23 ABL08194	Drosophila melanog
39	107.8	3.1	5664	23 ABL20330	Drosophila melanog
40	107.2	3.1	1920	24 ABK51681	DNA encoding human
41	107.2	3.1	2340	24 AAD22009	Human sitosterolae
42	107.2	3.1	2516	24 ABK51682	Human ABCG5 CDNA s
43	105.8	3.1	7869	23 ABL05134	Drosophila melanog
44	105	3.0	11206	21 AA247807	Vector for trappin
45	104.2	3.0	8486	24 ABN83687	Vector plasmid pCa

ALIGNMENTS

RESULT 1
AA294756
ID AA294756 standard: cDNA; 2258 BP.
AC AA294756;
XX
XX
DT 01-AUG-2000 (first entry)
XX Human ATP binding cassette cDNA fragment HuWhite2.
XX
XX HuWhite2; ATP binding cassette; human; cholesterol; lipid disorder;
KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
KW lupus erythematosus; diagnosis; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200018912-A2.
XX
XX 06-APR-2000.
XX
XX 21-SEP-1999; 99WO-EP06991.
XX
XX 25-SEP-1998; 98US-0101706.
XX (FARB) BAYER AG.
XX
XX Schmitz G, Klucken J;
XX
XX WPI; 2000-293151/25.
XX
XX Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders .

XX PS Claim 9; Page 135-136; 154pp; English.

XX CC The present sequence is that of human ATP binding cassette

CC (ABC) Huwhite2 cDNA, identified as a cholesterol-sensitive

CC gene fragment. The invention provides cholesterol-sensitive ABC

CC genes (see A294734-63). These genes, and polypeptides encoded by

CC them, can be used for diagnostic and therapeutic applications, and

CC for biochemical or cell-based assays to screen for pharmacologically

CC active modulator compounds useful for the treatment of lipid

CC disorders, atherosclerosis or other inflammatory diseases such as

XX psoriasis and lupus erythematosus.

XX SQ Sequence 2258 BP; 421 A; 716 C; 595 G; 525 T; 1 other;

Query Match 63.9%; Score 2207.8; DB 21; Length 2258;

Best Local Similarity 98.9%; Pred. No. 0;

Matches 2234; Conservative 0; Mismatches 18; Indels 6; Gaps. 1;

OY 64 ATGGCCGTGACGCTGGAGGACGGGGGAAACCCCTGTGCTGACACGACCTGAAGAAG 123

DB 1 ATGGCCGTGACGCTGGAGGACGGGGGAAACCCCTGTGCTGACACGACCTGAAGAAG 60

124 GTGAGAACACATCATCTGAAGCCAGCGCTTCCACCTGCCCCAAGCGCTCAGCCGTG 183

DB 61 GTGAGAACACATCATCTGAAGCCAGCGCTTCCACCTGCCCCAAGCGCTCAGCCGTG 120

OY 184 GACATCGAGTTCGTGGAGCTGCTCTATTCCGTGGGGAGGGGCCCTGTCGGCCAAAGG 243

DB 121 GACATCGAGTTCGTGGAGCTGCTCTATTCCGTGGGGAGGGGCCCTGTCGGCCAAAGG 180

OY 244 GGTATTAAGACCCCTTCTCAAGTGTCTCTCAGGTAAATTCGCCGCCGGAGCTGATTGGC 303

DB 181 GGTATTAAGACCCCTTCTCAAGTGTCTCTCAGGTAAATTCGCCGCCGGAGCTGATTGGC 240

OY 304 ATCATGGGCCCTCAGGGGCTGCAAGTCTACATTCATGTAACATCTTTGGCAGGATACAGG 363

DB 241 ATCATGGGCCCTCAGGGGCTGCAAGTCTACATTCATGTAACATCTTTGGCAGGATACAGG 300

OY 364 GAGTCTGAATGAAGGGCAGATCCTGTTAATGAAGGCCACCGGAGCTGAGGACCTTC 423

DB 301 GAGTCTGAATGAAGGGCAGATCCTGTTAATGAAGGGCACCAGGAGCTGAGGACCTTC 360

OY 424 CGCAAGATGTCCTGCTACATCATCAAGATGACATGCTGTGCCGACCTCACCGTGTG 483

DB 361 CGCAAGATGTCCTGCTACATCATCAAGATGACATGCTGTGCCGACCTCACCGTGTG 420

OY 484 GAAGCCATGATGCTCTGCTAAGCTGAATCTTACTGAGAAATCCCGATGTGAAGAAACGAT 543

DB 421 GAAGCCATGATGCTCTGCTAAGCTGAATCTTACTGAGAAATCCCGATGTGAAGAAACGAT 480

OY 544 CTGCTGACAGATCTCTGAGGACCTGGGCTGATGCTGCTCCACAGGAGGACGACC 603

DB 481 CTGCTGACAGATCTCTGAGGACCTGGGCTGATGCTGCTCCACAGGAGGACGACC 540

OY 604 CTGCTCTCTGGGGGACAGGAAGCGCTGCGCCATCGCCCTGGAGCTGGTCAACACCCG 663

DB 541 CTGCTCTCTGGGGGACAGGAAGCGCTGCGCCATCGCCCTGGAGCTGGTCAACACCCG 600

OY 664 CCGTCTGATGTTTGTGATGAGCCACAGGCTGCTGATAGCGGCTCTGTTTCCAAGTG 723

DB 601 CCGTCTGATGTTTGTGATGAGCCACAGGCTGCTGATAGCGGCTCTGTTTCCAAGTG 660

OY 724 GTGTCCTCATGAAGTCCCTGGACAGGGGGCGGTACCATCATCTGCACCATCCACAG 783

DB 561 GTGTCCTCATGAAGTCCCTGGACAGGGGGCGGTACCATCATCTGCACCATCCACAG 720

OY 784 CCCAGTCCCAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTGAGCCAGGGTCACTGC 843

DB 721 CCCAGTCCCAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTGAGCCAGGGTCACTGC 780

OY 844 ATCTTCAAGAGGTGCTCACCACCTGATCCCTATCTTAAGGGACCTCGGCTTGCAATGC 903

DB 781 ATCTTAAAGCGGTGGTCAACACCTGATCCCTATCTATAAGGAGCTCGGCTTGCAATGC 840

OY 904 CCCACCTACCAACAACCGGCTGACTTATCATCATCAG-----GTGGGCTCTGGCGAGTAT 957

DB 841 CCCACCTACCAACAACCGGCTGACTTCACTAGTGGGGGTCTGTGCTCTGGCGAGTAT 900

OY 958 GGAGACCTGAACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTGTGGCTATGGCTCAG 1017

DB 901 GGACACCTGAACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTGTGGCTATGGCTCAG 960

OY 1018 AAGAAGAGCAGCTCTGAGAAAGAGAGTCCCTGCCCATATGCTCTCTGCTCCCGAA 1077

DB 961 AAGAAGAGCAGCCTCTGAGAAAGAGAGTCCCTGCCCATATGCTCTCTGCTCCCGAA 1020

OY 1078 GTGGATCCCATTTGAAGCCACACCTTTGCCACAGACCCCTCACACAGTTCTGCATCCTC 1137

DB 1021 GTGGATCCCATTTGAAGCCACACCTTTGCCACAGACCCCTCACACAGTTCTGCATCCTC 1080

OY 1138 TTCAAGAGGACCTTCTCTGCTCCATCTCAGGACACGGTCTCTGACCCACCTACGCTTCATG 1197

DB 1081 TTCAAGAGGACCTTCTCTGCTCCATCTCAGGACACGGTCTCTGACCCACCTACGCTTCATG 1140

OY 1198 TCCACAGTGGTTATTTGGGGTCTCATCGGCTCTCTCTACCTGTCATATTTGGCGAGATGCC 1257

DB 1141 TCCACAGTGGTTATTTGGGGTCTCATCGGCTCTCTCTACCTGTCATATTTGGCGAGATGCC 1200

OY 1258 AGCAAGTCTTCAACAACACCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1317

DB 1201 AGCAAGTCTTCAACAACACCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260

OY 1318 GCCCTCATGCAACTGTCTCACCTTCCCTTTAGAGATGGCGTCTTCTCATGAGGAGACAC 1377

DB 1261 GCCCTCATGCAACTGTCTCACCTTCCCTTTAGAGATGGCGTCTTCTCATGAGGAGACAC 1320

OY 1378 CTCAACTACTGTTTACAGCCCTCAAAAGCGTATTTACCTGGCCAGACCATGGCTGACGTGCC 1437

DB 1321 CTCAACTACTGTTTACAGCCCTCAAAAGCGTATTTACCTGGCCAGACCATGGCTGACGTGCC 1380

OY 1438 TTTTCAGTGTGTGTCGGGTGCTACTGAGGATTTGTTACTGATGAGGAGGAGGAGGAGGAG 1497

DB 1381 TTTTCAGTGTGTGTCGGGTGCTACTGAGGATTTGTTACTGATGAGGAGGAGGAGGAGGAG 1440

OY 1498 GCTGAGACGAGCCCTTCT 1557

DB 1441 GCTGAGACGAGCCCTTCT 1500

OY 1558 CAATCTTTGGGGTGTGATGCGGAGCTGTCTTCAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 1617

DB 1501 CAATCTTTGGGGTGTGATGCGGAGCTGTCTTCAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560

OY 1618 GGCCAGTTACCGCATCCCTGCT 1677

DB 1561 GGCCAGTTACCGCATCCCTGCT 1620

OY 1678 ATCCCCACTTACCTTGAATGAGCT 1737

DB 1621 ATCCCCACTTACCTTGAATGAGCT 1680

OY 1738 GTGATCTGAGGATCTATGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1797

DB 1681 GTGATCTGAGGATCTATGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740

OY 1798 CCGTTCGGGAGGACAGAGATCTCTCCGAGCGTGGATGTGGAGGATGCCAAGCTCTAC 1857

DB 1741 CCGTTCGGGAGGACAGAGATCTCTCCGAGCGTGGATGTGGAGGATGCCAAGCTCTAC 1800

OY 1858 ATGAGTCTCTGTTTGGGCT 1917

DB 1801 ATGAGTCTCTGTTTGGGCT 1860

OY 1918 CTGGGTTTACCGGTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1977

DB 1861 CTGGGTTTACCGGTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1920

Db 901 ACCTACACACACCGGCTGACTCGTATCGAGTGGCCTCTGGCGAGTATGGAGCCTG 960
QY 967 AACCCCATGTTGTCAGGGCTGTCAGAAATGGCGTGTGGCTATGGCTGAGAAAGAGC 1026
Db 961 AACCCCATGTTGTCAGGGCTGTCAGAAATGGCGTGTGGCTATGGCTGAGAAAGAGC 1020
QY 1027 AGCCCTGAGAGAAAGAGGTCCTTGGCCCATGCTCTCTGTCCTCGGAAGTGGATCCC 1086
Db 1021 AGCCCTGAGAGAAAGAGGTCCTTGGCCCATGCTCTCTGTCCTCGGAAGTGGATCCC 1080
QY 1087 ATTGAAGCCACACCTTTGGCCACAGACCCCTCACAGAGTCTGCATCTCTTCAAGAGG 1146
Db 1081 ATTGAAGCCACACCTTTGGCCACAGACCCCTCACAGAGTCTGCATCTCTTCAAGAGG 1140
QY 1147 ACCTTCTGTCATCTCAGGACACGCTCTGACCCACCTACGGTTCATGTCTCCACGTC 1206
Db 1141 ACCTTCTGTCATCTCAGGACACGCTCTGACCCACCTACGGTTCATGTCTCCACGTC 1200
QY 1207 GTTATTGGCGTGTCTCATCGGCTCTCTACCTGCATATTGGCGACGATGCCAAGGTC 1266
Db 1201 GTTATTGGCGTGTCTCATCGGCTCTCTACCTGCATATTGGCGACGATGCCAAGGTC 1260
QY 1267 TTCAACAAACCGGCTGCTCTCTCTCTCATGCTGTCTCTCATGTTGCGCGCCCTCATG 1326
Db 1361 TTCAACAAACCGGCTGCTCTCTCTCTCATGCTGTCTCTCATGTTGCGCGCCCTCATG 1320
QY 1327 CCAACTGTGTACCTTCCCTTAGAGATGCGCTCTCTCATGAGGGAGCACCCTCACTAC 1386
Db 1321 CCAACTGTGTACCTTCCCTTAGAGATGCGCTCTCTCATGAGGGAGCACCCTCACTAC 1380
QY 1387 TGTGTACAGCCTCAAGCGTATTACCTGGCCAAAGACCATGGCTGACGTGCGCTTTCAGGTG 1446
Db 1381 TGTGTACAGCCTCAAGCGTATTACCTGGCCAAAGACCATGGCTGACGTGCGCTTTCAGGTG 1440
QY 1447 GTGTGTCGGTGTCTACTGACGATGTGTACTGATGACGGCCAGCCGCTGAGACC 1506
Db 1441 GTGTGTCGGTGTCTACTGACGATGTGTACTGATGACGGCCAGCCGCTGAGACC 1500
QY 1507 AGCCGCTTCTGCTCTTCTCAGCCCTGGCCACCGCCCTTGTGGGCCCAATCTTTG 1566
Db 1501 AGCCGCTTCTGCTCTTCTCAGCCCTGGCCACCGCCCTTGTGGGCCCAATCTTTG 1560
QY 1567 GGGCTGTGATGGAGTGTCTTCCAACTCCCTACAGGTGGCCCACTTTTGTGGGCCAGTT 1626
Db 1561 GGGCTGTGATGGAGTGTCTTCCAACTCCCTACAGGTGGCCCACTTTTGTGGGCCAGTT 1620
QY 1627 ACCGCCATCCCTGCTCTTCTGTCGGCTTCTTGTGACGTTCAAGACCATCCCACT 1686
Db 1621 ACCGCCATCCCTGCTCTTCTGTCGGCTTCTTGTGACGTTCAAGACCATCCCACT 1680
QY 1687 TACCTGCAATGGAGTGTCTTCTCTATGTGAGGTGGCCCACTTTTGTGGGCCAGTT 1746
Db 1681 TACCTGCAATGGAGTGTCTTCTCTATGTGAGGTGGCCCACTTTTGTGGGCCAGTT 1740
QY 1747 ACATCTATGGCATGGAGGAGGAGACCTGACATGTTTAGAGAACGCTGCCGTTCCGG 1806
Db 1741 ACATCTATGGCATGGAGGAGGAGACCTGACATGTTTAGAGAACGCTGCCGTTCCGG 1800
QY 1807 GAGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTC 1866
Db 1801 GAGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTC 1860
QY 1867 CTGGTCTGGGCATCTTCTTCTAGCCCTGCGGCTGCTGGCCCTACCTTGTGCTGCTTAC 1926
Db 1861 CTGGTCTGGGCATCTTCTTCTAGCCCTGCGGCTGCTGGCCCTACCTTGTGCTGCTTAC 1920
QY 1927 CGGGTCAAGTCAGAGAGATAG 1947
Db 1921 CGGGTCAAGTCAGAGAGATAG 1941

AAK51722
ID AAK51722 standard; cDNA; 3376 BP.
XX AC AAK51722;
XX AC AAK51722;
XX DT 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 267.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
DR P-PSDB; AAM78589.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 1; Page 1194-1198; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 3376 BP; 753 A; 945 C; 896 G; 782 T; 0 other;
SQ

Query Match 30.7%; Score 1062.4; DB 22; Length 3376;
Best Local Similarity 81.5%; Pred. No. 1.9e-254;
Matches 1389; Conservative 0; Mismatches 21; Indels 294; Gaps 2;
QY 227 CCTGCTGGCCGCAAGGGGTTATAGACCCCTTCTCAAGTGCCTCTCAGTAAATTCGTC 286
Db 38 CCGGAGGTCACACAGCTGTTATAGACCCCTTCTCAAGTGCCTCTCAGTAAATTCGTC 97
QY 287 GCGGGGAGCTGATTGGCATCATGGGCCCTCTCAGGGCTGGCAAGTCTACATTCATGAACA 346
Db 98 GCGGGGAGCTGATTGGCATCATGGGCCCTCTCAGGGCTGGCAAGTCTACATTCATGAACA 157

347 TCTTGGCAGGATACAGGAGTCTGGAATGAAGGGCAGATCCTGTTAATGAAGGCCAC 406
158 TCTTGGCAGGATACAGGAGTCTGGAATGAAGGGCAGATCCTGTTAATGAAGGCCAC 217
407 GGGAGCTGAGGACCTTCGCAAGATGCTCTGTACATCATGCAAGATGACATGCTGCTGC 466
218 GGGAGCTGAGGACCTTCGCAAGATGCTCTGTACATCATGCAAGATGACATGCTGCTGC 277
467 CCGACTCAGCGTGTGGAAGCATTGCTCTGCTAACTGAATCTTACTGAGAATC 526
278 CCGACTCAGCGTGTGGAAGCATTGCTCTGCTAACTGAATCTTACTGAGAATC 337
527 CCGATGTGAAAACGATCTGTGACAGAGATCCTGACGGCATTGGCCCTGATGCTGCT 586
338 AGGAGGTGAAGAGGAGTGTGTGACAGAGATCCTGACGGCATTGGCCCTGATGCTGCTGCT 397
587 CCCACAGGAGACCGCTGCTCTGCGGGCAGAGGAGCGTCTGCGCATCGCCCTGG 646
398 CCCACAGGAGACCGCTGCTCTGCGGGCAGAGGAGCGTCTGCGCATCGCCCTGG 457
647 AGCTGGTCAACAAACCGCTGCTCTGCTGATGAGCCACAGTGGTCTGATAGCG 706
458 AGCTGGTCAACAAACCGCTGCTCTGCTGATGAGCCACAGTGGTCTGATAGCG 517
707 CCTCTGTTTCCAAAGTGTGCTCTGATGAGTCCCTGGCAGAGGGGGCCGTACATCA 766
518 CCTCTGTTTCCAAAGTGTGCTCTGATGAGTCCCTGGCAGAGGGGGCCGTACATCA 577
767 TCTGACCATCCACAGCCAGTGTGCAAGCTCTGATGAGTCTGATGAGTCTGATGAGT 826
578 TCTGACCATCCACAGCCAGTGTGCAAGCTCTGATGAGTCTGATGAGTCTGATGAGT 625
827 TGAGCCAGGTCAGTGCATCTTCAAGAGGAGTGTGCAAGCTGTGATGAGTCTGATGAGT 886
626 -----AGTGCATCTTCAAGGGTGTGATGAGTCTGATGAGTCTGATGAGT 673
887 GACTCGGCTGATGTCGCTTCAAGAGGAGTGTGCAAGCTGTGATGAGTCTGATGAGT 946
674 GACTCGGCTGATGTCGCTTCAAGAGGAGTGTGCAAGCTGTGATGAGTCTGATGAGT 733
947 CTGGGAGTATGAGACCTGAAACCCATGTTTTCAGGCTGTGAGATGAGGCTGCGG 1006
734 CTGGGAGTATGAGACCTGAAACCCATGTTTTCAGGCTGTGAGATGAGGCTGCGG 793
1007 CTATGGCTGAGAAGAGCAGCCTTGAAGAAGAGTCTGCTGCTGCTGCTGCTGCTGCT 1066
794 CTATGGCTGAGAAGAGCAGCCTTGAAGAAGAGTCTGCTGCTGCTGCTGCTGCTGCT 853
1067 GTCTCCGGAAGTGGATCCATTGAAGCCACACCTTTGGCCACAGCAGCCTTCAACAGT 1126
854 GTCTCCGGAAGTGGATCCATTGAAGCCACACCTTTGGCCACAGCAGCCTTCAACAGT 913
1127 TCTGCATCCTCTTCAAGAGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1186
914 TCTGCATCCTCTTCAAGAGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
1187 TACGGTTCATGTCCTCCACGTTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1246
960 ----- 959
1247 GCGAGATGCCAGAGGCTTCAACACACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1306
960 ----- 959
1307 TATGTTCCGCGCTCATGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1366
960 ----- 959
1367 TGAGGAGCACCTCACTACTGTTACAGCTTCAAGCGTATTACCTGGCCAGAGCATGG 1426
960 ----- 959

1427 CTGAGCTGCCCTTTCAGGTGTTGTCGCGGTGTTCTACTGACAGCATTTGTACTGATGA 1486
960 -----GGTGGTGTGTCGCGGTGTTCTACTGACAGCATTTGTACTGATGA 1003
1487 CGGGCCAGCCGCTGAGACCAAGCCTTCCCTGCTCTTCTCAGCCCTGCGCCACGCCACCG 1546
1004 CGGGCCAGCCGCTGAGACCAAGCCTTCCCTGCTCTTCTCAGCCCTGCGCCACGCCACCG 1063
1547 CCTTGGTGGCCCAATCTTTGGGCTGCTGATCGAGCTGCTTCCAACTCCCTACAGGTGG 1606
1064 CCTTGGTGGCCCAATCTTTGGGCTGCTGATCGAGCTGCTTCCAACTCCCTACAGGTGG 1123
1607 CCACCTTTTGTGGGCGCAAGTTACCGCAATCCCTGCTCTTGTGTTCCGGCTTCTTTGTGA 1666
1124 CCACCTTTTGTGGGCGCAAGTTACCGCAATCCCTGCTCTTGTGTTCCGGCTTCTTTGTGA 1183
1667 GCTTCAAGACCATCCCACTTACCTGCAATGAGCTCCTATCTCTCTATGTCAGGTATG 1726
1184 GCTTCAAGACCATCCCACTTACCTGCAATGAGCTCCTATCTCTCTATGTCAGGTATG 1243
1727 GCTTGAAGGCTGATCTCTGACGATCTATGCGATGAGCGAGGAGACCTTGACATGTTTATG 1786
1244 GCTTGAAGGCTGATCTCTGACGATCTATGCGATGAGCGAGGAGACCTTGACATGTTTATG 1303
1787 AGGAACGCTGCCGTTTCCGGGAGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGGATG 1846
1304 AGGAACGCTGCCGTTTCCGGGAGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGGATG 1363
1847 CCAAGCTCTACATGGAATCTTCCCTGCTTGGGCTGCTCTCTCTAGCCCTGCGCTGCTGG 1906
1364 CCAAGCTCTACATGGAATCTTCCCTGCTTGGGCTGCTCTCTCTAGCCCTGCGCTGCTGG 1423
1907 CCTACTTGTGCTGCTTACCGGG 1930
1424 CCTACTTGTGCTGCTTACCGGG 1447
RESULT 4
AAH90113
ID AAH90113 standard; cDNA; 3376 BP.
XX
AC AAH90113;
XX
DT 01-OCT-2001 (first entry)
XX
Human bone marrow cDNA, SEQ ID NO: 470.
XX
Human; bone marrow; antineoplastic; cytostatic; neuroprotective;
antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
immunosuppressive; gene therapy; cytokine cell proliferation;
cell differentiation modulator; immune disorder; infection; cancer;
human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.
OS Homo sapiens.
XX
PN WO200153453-A2.
XX
PD 26-JUL-2001.
XX
PF 23-DEC-2000; 2000WO-US34960.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 30-NOV-2000; 2000US-0250583.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;

Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Drmanac RT;
WPI: 2001-488707/53.
P-PSDB; RAA00994.

Novel bone-marrow-expressed polynucleotides and polypeptides, useful for treating e.g. cancer and immune deficiency disorders -
Claim 1; Page 592-596; 648pp; English.

The present sequence is one of 251 novel human polynucleotides expressed in the bone marrow. The polynucleotide and the polypeptide encoded by it are useful in the treatment of various immune deficiencies and disorders. The deficiencies and disorders may be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal infection, or may result from an autoimmune disorder, a coagulation disorder (e.g. haemophilia), inhibition of tumour cell proliferation, suppression of an inflammatory response or treatment of a nervous system disorder such as Alzheimer's disease. Detection of the presence or increased expression of the polynucleotide or the protein it encodes is useful for the diagnosis and/or prognosis of one or more types of cancer. The polynucleotide and polypeptide can be used as nutritional sources or supplements and in the screening of chemical compounds as potential drugs.

Sequence 3376 BP; 753 A; 945 C; 896 G; 782 T; 0 other;

Query Match 30.7%; Score 1062.4; DB 22; Length 3376;
Best Local Similarity 81.5%; Pred. NO. 1.9e-254;
Matches 1389; Conservative 0; Mismatches 21; Indels 294; Gaps 2

QY	227	CCTCTCTGGCGCAAAAGGGTTTATAAGACCCCTTCTCAAGTGCTCTCAGGTAAATTTCTGCC	286
DB	38	CCCAGGTCACACAGCTGGTTATAAGACCCCTTCTCAAGTGCTCTCAGGTAAATTTCTGCC	97
QY	287	GCCGGGAGCTGATTGGCATCATGGGCCCCCTCAGGGGCTGGCAAGTCTACATTCATGAACA	346
DB	98	GCCGGGAGCTGATTGGCATCATGGGCCCCCTCAGGGGCTGGCAAGTCTACATTCATGAACA	157
QY	347	TC TTGGCAGGATACAGGAGTCTGGAAATGAAGGGCAGATCCTGGTTAATGGAAGGCCAC	406
DB	158	TC TTGGCAGGATACAGGAGTCTGGAAATGAAGGGCAGATCCTGGTTAATGGAAGGCCAC	217
QY	407	GGGAGCTGAGGACCTTCGGCAAGATGTCTGCTACATCATGCAAGTACATGCTGCTGTC	466
DB	218	GGGAGCTGAGGACCTTCGGCAAGATGTCTGCTACATCATGCAAGTACATGCTGCTGTC	277
QY	467	CGCACCTCAGGTTGGTGGAGCCATGATGGTCTCTGCTTAACCTGAACTTACTGAGAATC	526
DB	278	CGCACCTCAGGTTGGTGGAGCCATGATGGTCTCTGCTTAACCTGAACTGAGTGAAGAAGC	337
QY	527	CCGATGTGAAAAAGCATCTCGTGACAGAGATCCTGACGGCACTGGGCCCTGATGCTGCTGT	586
DB	338	AGGAGGTGGAAGAGGAGCTGCTGACAGAGATCCTGACGGCACTGGGCCCTGATGCTGCTGT	397
QY	587	CCACACAGGACAGCCCTGCTCTCTGCGGGCAGAGAAACGCTCTGCGCCATCGCCCTGG	646
DB	398	CCACACAGGACAGCCCTGCTCTCTGCGGGCAGAGAAACGCTCTGCGCCATCGCCCTGG	457
QY	647	AGCTGGTCAACAACCCGCTGTCTATGTTCTTTTGATGACCCACCAGTGGTCTGGATAGCG	706
DB	458	AGCTGGTCAACAACCCGCTGTCTATGTTCTTTTGATGACCCACCAGTGGTCTGGATAGCG	517
QY	707	CCTCTGTTTCCAAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGGGGGCGGTACCATCA	766
DB	518	CCTCTGTTTCCAAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGGGGGCGGTACCATCA	577
QY	767	TC TGCACCATCCACAGCCAGTGCCTGCTTTTGAGATGTTTGACAAAGCTCTACATCC	826
DB	578	TC TGCACCATCCACAGCCAGTGCCTGCTTTTGAGATGTTTGACAAAGCTCTACATCC	826
QY	827	TGAGCCAGGTCAGTGCATCTTCAAGGAGTGGTGCACCAACCTGATCCCTTATCTAAAGG	886

Db 1424 CCTACCTTGTGCTGCTTACCGGG 1447

RESULT 5
ABI99363
ID ABI99363 standard; cDNA; 3586 BP.
XX
AC ABI99363;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:254.
XX
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS Mus musculus.
XX
W0200188188-A2.
XX
PD 22-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-JP04192.
XX
PR 18-MAY-2000; 2000JP-0145977.
XX
PA (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
DR WPI; 2002-034733/04.
DR P-PSDB; ABB57112.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
XX
XX Claim 2; Page 714-719; 2690pp; English.
XX
CC The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC the expression levels of particular genes (ABI99202 to ABI99912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischaemic condition-improving
CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 3586 BP; 810 A; 979 C; 920 G; 877 T; 0 other;

Query Match 30.2%; Score 1042.2; DB 24; Length 3586;
Best Local Similarity 72.3%; Pred. No. 2.2e-249;
Matches 1385; Conservative 0; Mismatches 518; Indels 12; Gaps 2;

QY 79 GAGACGGGGGGAACCCCTGTGCTGACCAACGCTCAGCCGCTGGACATCGAGTTCGT 198
Db 200 GTGGATGAGTTGAGACAGACCTGCTCAATCGGCACCTGAAGAAGGTGGACAACATTC 259
QY 139 ACTGAAGCCAGCGCTTCTCCACCTGCCCAAGGCTCAGCCGCTGGACATCGAGTTCGT 198
Db 260 ACAGAGCCAGCGCTTCTCCCTTCCCGGAGGCGCGCTGAACATCGGAATTCAG 319
QY 199 GAGCTGTCTTATTCGCTGGGGGGGCGCTGTGCGGCAAAAGGGGTTATAAGACCCCTT 258
Db 320 GACCTTCTACTCTGTACCCGAGGGGCGCTGTGGAGAGAAAGATACAGACCCCTT 379

QY 259 CTCAAGTGCCTCTCAGGTAAATTTCTGCCGCCGGAGAGTGAATGGCATCATGGGCCCCCTCA 318
Db 380 TTCAAAAGGGATCTCTGGGAAATTTCAACAGTGGAGAGATGCTGGCCATCATATGGTTCCTT 439
QY 319 GGGCTGGCAAGTCTACATTCATGAACATCTTGCAGGATACAGGGAGTCTGGAATCAAG 378
Db 440 GGAGCTGGGAAGTCCACACTCATGATATTTCTGGCAGGATACAGGGAGATGGCATGAAA 499
QY 379 GGGCAGATCCTGTTAATGAAGGCCACGGGAGCTCAGGACCTTTCGCCAAGATGTCCTGC 438
Db 500 GGGCAGTCTTATCAATGAATGCCCGGAGCTGCGCTCTCCGGAAGTCTCCTGC 559
QY 439 TACATCATGCAAGATGACATGCTGCTGCCACCTCACGGTGTGGAGGATCATGATGGTC 498
Db 560 TACATCATGCAAGATGACATGCTGCTGCCACCTCACGGTGTGGAGGATCATGATGGTC 619
QY 499 TCTGCTAACCTGAATCTTACTGACATCCCGATCTGAAAACGATCTGTCGTGACAGATC 558
Db 620 TCCGCGCATCTGAAGCTGACAGAGAGGATGAAGGACAGCGGGAGATGGTCAAGAGATC 679
QY 559 CTGACGGCACTGGGCGCTGATGCTGCCACACAGAGAGACAGCCCTGCTCTCTGCGCGG 618
Db 680 CTGACAGCCCTGGCTGCTGCCCTGTGCCAACACACAGCAGCGGGAGCCCTCTCAGGCGG 739
QY 619 CAGAGGAAGCTGTGGCCATGCGCCCTGGAGCTGGTCAACAACCGCGCTGTCATGTTCTT 678
Db 740 CAGCGGAACGCTTGGCCATTTGCCCTGGAGCTGGTCAACAACCGCGCTGTCATGTTCTT 799
QY 679 GATGAGCCCACTGCTGCTGATAGCGCTCTGTTTCCCAAGTGGTGTCCCTCATGAAG 738
Db 800 GATGAGCCCACTGCTGCTGATAGCGCTCTGTTTCCCAAGTGGTGTCCCTCATGAAG 859
QY 739 TCCCTGGCAGAGGGGCGCTGATCATCTGCACCATCCACAGCCAGTGCACAGCTC 798
Db 860 GGACTGCCAGGGTGGCGCTCCATCTCTGTACCATCCACAGCCAGTGCACAGCTC 919
QY 799 TTTGAGATGTTTGAACAAGCTTACATCTGAGCCAGGCTCAGTGCATCTTCAAAGGAGT 858
Db 920 TTTGAGCTTTTGACAGCTTATGCTAAGTCAAGGACATGCTGATACAGGGAAG 979
QY 859 GTCACCAACCTGATCCCTATCTAAAGGAGTGGCTTGCATTTGCCCCACCTACCAAC 918
Db 980 GTCTCCAACTCTGTCGCGTATCTGAGGATCTGGGCTGTAACCTGCTACCTACCAAC 1039
QY 919 CCGCTGACTTCACTCATGAGAGTGGCTCTGGGAGTATGAGACCTGAAACCCCATGTTG 978
Db 1040 CCAGCAGCTTGTCTATGGAAGTGGCATCAGGGAGTACGGGATCAGAACAGTGCCTG 1099
QY 979 TTAGGCTGTGCAAGATGGCTGTGCGCTATGGCTGAGAGAGAGAGAGCCCTGAGAG 1038
Db 1100 GTGAGAGCCGTGGCAGAGGCGCATGTGTAGCGCTGACTATAAGAGAGACCTCGGGGCG 1159
QY 1039 AACGAGTCTCCCTGCCCATGCCCTCTGCTTCCCGAAGTGGATCCCATTTGAA---- 1093
Db 1160 ACCGATGTGAACCGGTTTCTTTGGCAGCGGCTCTGTAAGAGGACTCCGCTCCATGGAA 1219
QY 1094 ----GCCACACCTTTGGCAGCAGCCCTCACACAGTTCTGCATCTCTTCAAGAGGACC 1149
Db 1220 GGTGTCATAGCTTCTCGGCGAGCTGCTCACCCAGTTCCTGCATCTCTTCAAGAGGACC 1279
QY 1150 TTCTGTTCATCTCAGGAGACAGGCTCTGACCCACCTACGGTTCATGTGCCAGGCTGGT 1209
Db 1280 TTCTCAGCATCATGCGGAGCTCGGCTCTGACACATCTGCGAATTCAGCTCGCATTTGG 1339
QY 1210 ATTGGGCTGTCTATCGGCTCTCTACCTGCATATTGCGAGGATGCCAGAGTCTCTC 1269
Db 1340 ATGCGCTGTCTATTTGCGCTCTGCTACCTGGGATTTGGGATGAAGCCAGAGGCTCTT 1399
QY 1270 AACAAACCGGCTGCTCTTCTTCTCCATGCTGTTCCTCATGTTTTCGCGGCTCATGCCA 1329
Db 1400 AGCAACTCCGCTTCTCTTCTTCTCCATGCTGTTCCTCATGTTTGTGCTGCTCATGCC 1459
QY 1330 ACTGTGCTACCTTCCCTTTAGAGATGGCGTCTTTCATGAGGAGGAGACCTCAACTACTGG 1389

1460 ACTGTTCTGACCTTTCCCTCGGAGATGAGTGTCTTCTCCGGGAGCACCTGAACCTACTGG 1519
1390 TACAGCCTCAAGCGTATTACCTGGCCCAAGACCATGGCTGACGTGCGCTTTCAGGTGGT 1449
1520 TACAGCCTGAAGCGCTACTACCTGGGCAAGACCATGGCCGATGTCCTCTTTCAGATCATG 1579
1450 TGTCGGGTGCTCTACTGACGATTTGTCTGATGACGGCCAGCCCGCTGAGACCAAGC 1509
1580 TTCCCTGTGCGCTACTGACGATTCGTATCTGATGACGTCACGCGCTGGACGCTGTG 1639
1510 CGCTTCCTGCTCTCTCAGCCCTGGCCACCGCCACCGCTTGTGGCCCAATCTTTGGGG 1569
1640 CGTTTGTGCTGTCTGCTGCTGGTACCATGACATGACATGCTGTGGCCAGTCTCTAGGA 1699
1570 CTGCTGATGCGAGCTGCTCCCACTCCCTACAGTGGCCACTTTTGGGGCCCACTTACC 1629
1700 CTACTGATGAGCTGATCCATCCATCCCTCGAGCTTGCAGATTTGGGTGCGGTGACA 1759
1630 GCCATCCCTGCTCTTCTTCTCCGGCTTCTTGTCTGAGCTTCAAGACCATCCCACTTAC 1689
1760 GCCATCCCTGCTCTTCTTCTCCGGATTTCTTGTGACGCTTTGACACCATCCCACTTAC 1819
1690 CTGCAATGAGCTCTCTATCTCTCTATGTCAGGTATGCTTGTGAGGTGTGATCTCTGAGC 1749
1820 CTCAGTGGATGCTCTACATCTCTATGTCAGATACGCTTTGAGGGGTCTATCTCTGTC 1879
1750 ATCTATGCTGATGAGCGAGGACCTGACATG--TTTACAGGACGCTCCCGTCCGG 1806
1880 ATCTAGGCTTGGACCGAGAGACCTGCTGCTGACATCCGCGAGACATGCCACTTCAG 1939
1807 GAGCCACAGACATCTCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTC 1866
1940 AAGTCAGAGCCATCTCTGAGGAGCTGGACGTGAGATCGAAGCTGTACCTGGATTTC 1999
1867 CTGGTCTGGGATCTCTTCTTCTAGCGCTCGGCTGCTGCGCTACCTTGTGCTGCGTTAC 1926
2000 ATGCTCTGGGATCTCTTCTTCTATCTGCTGCGCTCATCTGCGCTATTTGCTCTCAGATC 2059
1927 CGGGTCAAGTCAGAGATAGAGGCTTGCCCCAGCGCTGTACCCAGCGCCCTGCGAG 1981
2060 AAATCCGGCTGAGAGGTAAACCCGCTTCCAGCCAGCAGGACGAGCAAGCAG 2114

RESULT 6

AAZ94747

ID AAZ94747 standard; cDNA; 2930 BP.

XX AC AAZ94747;

XX 01-AUG-2000 (first entry)

XX Human ATP binding cassette ABCG1 (ABCG1) cDNA.

XX ABCG1; ABC8; ATP binding cassette; human; phagocyte; cholesterol;
KW phospholipid; transporter; inflammation; atherosclerosis;
KW lipid disorder; dyslipidemia; psoriasis; lupus erythematosus;
KW diagnosis; gene therapy; chromosome 21q22.3; ss.

XX Homo sapiens.

XX W0200018912-A2.

XX 06-APR-2000.

XX 21-SEP-1999; 99WO-EP06991.

XX 25-SEP-1998; 98US-0101706.

XX (FARB) BAYER AG.

XX Schmitz G, Klucken J;

XX

DR

XX

PT

PT

XX

PS

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

SQ

Query Match 28.2%; Score 972.8; DB 21; Length 2930;

Best Local Similarity 70.5%; Pred. No. 4e-232;

Matches 1351; Conservative 0; Mismatches 517; Indels 48; Gaps 2;

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

WPI; 2000-293151/25.
Adenosine triphosphate binding proteins useful for identifying agents for treating atherosclerosis and other inflammatory disorders -
Claim 9; Page 129-130; 154pp; English.

The present sequence is that of human cDNA encoding ATP binding cassette protein ABCG1 (ABC8). The cDNA was identified using a differential display method in which monocytes from peripheral blood were subjected to macrophage differentiation and cholesterol loading with acetylated low density lipoproteins and subsequent degrading with high density lipoprotein (HDL3) to identify cholesterol sensitive genes. The ABCG1 gene maps to human chromosome 21q22.3. ABCG1 was identified as a cholesterol sensitive switch. It is up-regulated by macrophage colony stimulating factor dependent phagocytic differentiation, and expression is massively induced by cholesterol loading and almost completely set back to differentiation dependent levels by HDL3. ABCG1 is the human homologue of the Drosophila white gene. Sequencing of the promoter (see AA294784) revealed important transcription factor binding sites relevant for phagocytic differentiation and lipid sensitivity. ABCG1 was also identified as a cholesterol transporter and modulator of choline-containing phospholipids (phosphatidylcholine, sphingomyelin). The invention provides cholesterol-sensitive ABC genes (see AA294734-63) that can be used for diagnostic and therapeutic applications, and for biochemical or cell-based assays to screen for pharmacologically active compounds useful for the treatment of lipid disorders, atherosclerosis or other inflammatory diseases such as psoriasis and lupus erythematosus.

Sequence 2930 BP; 665 A; 770 C; 780 G; 715 T; 0 other;

81 GGAGCGGGGGAACCCCTGCTGTGCTGACCGACCTGGAAGAGGTGAGAACACATCAC 140
141 GGAGGCCATGACGACGACCTGCTGATGACATCTGAAAAAGTAGATATACCTCAC 200
141 TGAAGCCCAAGCGCTTCTCCACCTGCCCAAGCGCTCAGCGTGACATGAGTTCGTGGA. 200
201 GGAAGCCCAAGCGCTTCTCCCTCTGCTCGGAGGAGCTGTGAACATTAATTCAGGA 260
201 GCTGTCTATTCCGTGGGGAGGGCCCTCTCTGGCGAAGAGGGTTATAGACCTCTCT 260
261 CCTTCTCTATTCTGCTGAGGAGCCCTGCTGAGGAGCCCTGCTGAGGAGGATACAGACCTCTCT 320
261 CAAGTGCCTCTCAGGTAATTTCTCCGCGGAGCTGATGGCATCATGGCCCTCAGG 320
321 GAAAGGAATTTCCGGAAGTTCAATAGTGTGAGTTGGTGGCCATTATGGTCTCTCCCG 380
321 GGCTGGCAAGTCTACATTTATGAAACATCTTTGGCAGGATACAGGAGTCTGGAATGAAGG 380
381 GGCCGGGAAGTCCACCTGATGAACATCTCTGCTGGATACAGGAGCGGATGAAGG 440
381 GCAGATCTCTGGTTAATGGAAGCCACGGAGCTGAGGACCTTCCGGAAGATGCTCTGCTA 440
441 GGCCGCTCTCATCAAGGGCTGCGCGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA 500
441 CATCATGCAAGATGACATGCTGCTGCGGACCTCAGGCTGCTGCTGCTGCTGCTGCTGCTC 500
501 CATCATGCAAGATGACATGCTGCTGCGGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560
501 TGCTAACTCTGAATCTTACTGGAATCCCGATGTGAAAAACGATCTGCTGACAGATGCTCT 560
561 GGCACATCTGAAGCTTCAGGAGAGGATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 620
561 GACGGCACTGGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 620
621 GACAGCGCTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680

02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX
XX (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
XX
XX Claim 1: SEQ ID 1658: 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX carcinoma, papillary carcinoma and Wilms' tumour.
XX
XX Sequence 2930 BP: 665 A; 770 C; 780 G; 715 T; 0 other: 70

Query Match	28.2%	Score 972.8;	DB 24;	Length 2930;
Best Local Similarity	70.5%;	Prod. No. 4e-232;		
Matches 1351;	Conservative	0;	Mismatches 517;	Indels 48; Gaps 2;
81	GGACGGGGGGACCCCTGCTGTGACCACGCGACCTGAAGAAGTGGAGAACCATCAC	140		
141	GGAGGCCACTGAGACGGACCTGCTGATGACATCTGAAAAAGTAGATAATAACCTCAC	200		
141	TGAAGCCACGCGCTTCTCCACCTGCCCCAGCGCTCAGCCGTCGACATCGAGTTCGTGGA	200		
201	GGAAGCCACGCGCTTCTCCCTGCTGCGAGGGCAGCTCTGAACATTTGAATTCAGGGA	260		
201	GCTGTCTTATTCGTCGGGAGGGCCCTCTCTGGCGCAAAAGGGTTATAGACCCCTCT	260		
261	CTTTCTCTATTCGGTTCCTGAAGGACCCCTGGTGGAGGAAGAAGGATACAAAGACCCCTCT	320		
261	CAAGTGCCTCTCAGGTAAATTCTGCGCCCGGAGCTGATTGGCATCATGGCCCTTCAGG	320		
321	GAAGGAATTCGGGAAGTTCAAATAGTGTGAGTTGGTGCCCATATGGTCTCTCCGG	380		
321	GCGTCCAACTACATTCATGAACATCTTTGGCAGGATACAGGGAGTCTGGAATGAAGG	380		
381	GCCCGGGAATCCACGCTGATGAACATCTCTGGTGGATACAGGGAGACGGCATGAAGG	440		
381	GCAGATCTCGTTAATGGAAGGCCAGGGAGCTGAGGACCTTCGCGAAGATGTCTCTGCTA	440		


```
Db 1158 GCCCTGTCGATGTACCAATCCAGCGGATTACATTATCGAGCTCGCTTTCGCGAGAGTA 1217
Qy 957 TGGAGACTGAACCCCATGTTTTCAGGGCTGTGCAGATGGGCTGTGCGCTATATGGCTGA 1016
Db 1218 TGGATACGACAGGTTCACACCCTTAAGCTCGCCACTGAGAACGGAGAGCTGCTGACTTG 1277
Qy 1017 GAAGAAGAGCAGCCCTGAGAGAACGAGGTCCTTCCCTCCATGCCCTCTTGTCTCCGGA 1076
Db 1278 GTTCCATAATCCCAAGTGGCGGTGTCGCGCTGAGGCTTGTATGAGAAAGTATCCATACC 1337
Qy 1077 AGTGGATCCCATTAAGACACACCTTTGCCACACGACCCCTCACAGTTCCTCATCCT 1136
Db 1338 AAAGAAACCAAGAGCCGCTCTCTGCGGAGACACACGCTACTCAAAACGAGCTCTGTGCT 1337
Qy 1137 CTTCAAGAGGAGCTTCTGCTCCATCTCTCAGGACACGCTCTGACCCACCTTACGTTTCAT 1196
Db 1398 TTTCCGACGAGGTTTCATCAAGCCAAACGACACGACCATGACGCACTTGAGATCGG 1457
Qy 1197 GTCCACGCTGTTATGGCGTCTCATGGCCCTCTCTACCTGCATATTTGGCGACGATGC 1256
Db 1458 AGTCAACATAGCAGTGGCGCTCTGTTCCGAGCTATGTACGATCACACGGGACGCTGAAG 1517
Qy 1257 CAGCAAGCTTTCACAAACACCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1316
Db 1518 ATCAGAGTGTCTGACAACTACAACTCTCTGTTTGGCCATCTGATGACCACTTCCATGAC 1577
Qy 1317 CGCCCTCATGCAACTGTGCTCACCTTCCCTTAGAGATGCGGTCTTCATGAGGGAGCA 1376
Db 1578 CACGATGATGTTGACTGTTCTGACTTTCCTCCCATGGACATCTCCATATTTGATCAAGGAC 1637
Qy 1377 CTTCAACTGCTGTTACAGCTCAAGGCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1436
Db 1638 CTTCAACTGCTGTTACAGCTCAAGGCTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1697
Qy 1437 CTTTCAGTGTGTCGCGGTGCTACTGACGATGCTGACGATGCTGACGCGGCGGAC 1496
Db 1698 CATATCTATCAATAGTTGTTCTTCTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1757
Qy 1497 CGCTGAGACACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1556
Db 1758 GATGAGTGGATCAGATCTTATGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1817
Qy 1557 CCAATCTTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1616
Db 1818 ACATAGTTTGGTGTGATGATTGG---TGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1874
Qy 1617 GGGCCCACTTACCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1676
Db 1875 GGGCCCACTTACCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1934
Qy 1677 CATCCCACTTACCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1736
Db 1935 TCTGCCAAGCTACTTAAGTGGGCGAGTCAATATCATATCATATCATATCATATCATAT 1994
Qy 1737 TGTGATCTGACGATCTATGCGATGAGGAGGAGACCTGACATGTTTATGAGGAAC---G 1793
Db 1995 CTTTATATCAGCCATTTATGCTTGGATGAGGATGCTTGGCTTGGAGGAGGCGCGCTA 2054
Qy 1794 CTGCGCTTCCGCGGAGCAGACGATCCTCCGAGCGCTGATGCTGGAGGATGCCAAGCT 1853
Db 2055 CTGCGCTTACAGGATCCAAAGAAATTTCTTAGAGGAATACACATGAGGGGTGATCAGTT 2114
Qy 1854 CTACATGAGCTTCTGCTTGGGATCTCTCTCTAGCCCTGCGGCTGCTGCGCTTACT 1913
Db 2115 CTGGAACGATGATGATGCGCTGGGAGTCAATCTCTGCTTTCGATTTGCTCTAGCT 2174
Qy 1914 TGTGCTGGTGTACCGGGTCAAGTCAGAGATAG 1947
Db 2175 GGTGCTGAAGGCCAAGATCAAGTCGATCCGATAG 2208
```

RESULT 10
ABL05663

```
ID ABL05663 standard; cDNA; 2283 BP.
XX AC ABL05663;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11471.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ss.
XX PN Drosophila melanogaster.
XX PD WO200171042-A2.
XX PF 27-SEP-2001.
XX PP 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX P-PSDB; ABB61560.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PS interactions -
XX PS Claim 1; SEQ ID NO 11471; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2283 BP; 586 A; 538 C; 493 G; 666 T; 0 other;
Query Match 11.3%; Score 391.8; DB 23; Length 2283;
Best Local Similarity 53.2%; Pred. No. 2.6e-87;
Matches 958; Conservative 0; Mismatches 792; Indels 51; Gaps 4;
Qy 155 TCTCCCACTGCCCAAGCGCTCAGCCGTGGACATCGAGTTCGTGGAGCTGCTTATTCG 214
Db 104 TCCAGCACTCCCAAGCGACCGGTGGACCTGGCTTTCACAACTGACGTACCGCG 163
Qy 215 TCGCGGAGGCGCTGCTGCGCCAAAGGGGTATAGACCCCTTCTCAAGTGCCTCTCAG 274
Db 164 TCAAGGAGGGAACCGGAG-----CAACCATCTGAGGGCGTCAGTG 206
Qy 275 GTAAATCTCGCGCGGAGCTGATTTGGCATATGAGGCGCTCGAGGCTGGCAAGTCA 334
Db 207 GACGCTCCGCTCTGGAGAACTCACCGCCATANTGGGACCATCGGCTGCTGGCAAGTA 266
Qy 335 CATTATGAACATCTTGGCAGGATACAGGGAGTCTGGAATGAAGGGGAGATCTTGGTTA 394
Db 267 CGCTGCTCAACATTTTGTCTGGCTACAAAACCTTCAGCATCGAAGGCGAGCTTACGATGA 326
Qy 395 ATGGAAGCCACGCGGAGCTGAGGACCTTCCCGCAAGATCTCTGCTACATCATGCAAGATG 454
Db 327 ACGGTGCGGAGCGCAACCTAAGCGCATTTTCGCAAGCTTTCTGCTACATCATGCGAGCA 386
```


CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AB57737-AB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 2639 BP; 661 A; 651 C; 650 G; 677 T; 0 other;

	Query Match	9.6%	Score 332.8;	DB 23;	Length 2639;
	Best Local Similarity	52.0%;	Pred. No. 1.4e-72;		
	Matches 877;	Conservative 0;	Mismatches 787;	Indels 22;	Gaps 5;
QY	233	GGCCAAAGGGTTATAGACCCCTCTCAAGTGCCTCTCAGGTAAATTTGCCCGCGG	292		
Db	910	GGCCACCAAGGATCCACTCCCATATACAGAGGCGTGGGGTCTTCAATCGGGAC	969		
QY	293	AGCTGATTTGGCATCATGGGCCCCCTCAGGGCTGCGAAGTCTACATTCATGAACATCTTGG	352		
Db	970	GACTTACGGCCATTTTGGGACCCCTCAGGAGCTGGAAGTCCACCTTGTCTTAATGCTTTGG	1029		
QY	353	CAGGATACAGGAGTCTGGAATGAGGGCGAGATCCTGTTAATGGAAGGCCACGGGAGC	412		
Db	1030	CGGGTTTAAACTCCAGGGAGTTAGGGCCAGTTCTGTTGAATGGAGCCCCCAGGGATA	1089		
QY	413	TCAGGACCTTCGCAAGATGTCCTGTATCATCATCAAGATGACATGCTGCTGCCGACC	472		
Db	1090	TAATGAGCTTTCCGGAATATGTCGCTATATGCTCAGAACTTTCTGATGCTAAATCTGC	1149		
QY	473	TCAGGCTTTGGAAGCCATGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	526		
Db	1150	TCACGGTTGAGGAACACTTGGGAGTTTCTACAGACCTGAAGATGCCAGTTCACAGCTG	1209		
QY	527	CGATGTCAAAACCATCTCTGACAGAGATCCCTGACGCGACTGGCGCTGATGCTGCT	586		
Db	1210	CTCAGAGAACAGAAATTTAGACGATATATCGACATCTGAGCTACAACTCCTGCC	1269		
QY	587	CCACACAGGACAGCCCTGCTCTGCGGCGAGAGGAGCTGCGCCATCGCCCTGG	646		
Db	1270	GCAGGACCTTGGTGAAGAATTTATCCGGAGCGGACACAGGCTTGTCCATTTGCCATTG	1329		
QY	647	AGCTGGTCAACACCCGCTGCTGATGTTCTTTGATGAGCCACCAAGTGGCTGGATAGCG	706		
Db	1330	AGTTGGTAAACCAATCCGCGCAATTTATGTTCTCGATGAGCCACCAAGCGGATTTGGATTG	1389		
QY	707	CCTCTGTTTCCAGTGGTGCCTCATGAAGTCCCTGCGACAGGGGCGCTACCATCA	766		
Db	1390	TGGGAGCTACAGGTGATTTGCCATCTGCGAGGACTGGCGACGATGGCAGGATAGTTG	1449		
QY	767	TCTGCACCATCCACAGCCCATGCGCAAGCTCTTTGAGATGTTTGACAAAGCTCTACATCC	826		
Db	1450	TTTGTGTTTCATCAGCGGGATCGCGACTCTTCCAACTCTTCATGATGCTCTGGTTT	1509		
QY	827	TGAGCCAGGTCAGTGCATCTTCAAGAGAGTGGTCAACCACTGATCCCTCTATCTAAAG	886		
Db	1510	TGGCTCAGCGAGAGTCTTTACGCTGGGGAACACGCGGAATGCTGCCACTTTCCGAC	1569		
QY	887	GACTCGGCTTGATTTGCCCATCCACACACCCGCTGACTTTCATCATCGAGTGGCT	946		
Db	1570	ATATCGGCACATCTGTCCACAGTACTACATCCCGGATTTCCGCTGGGAAGTGCA	1629		
QY	947	CTGGCGAGTATGAGACCTTGAAACCCCATGTTTTCAGGGCTGTGAGAAATGGGCTGTGCG	1006		
Db	1630	GTCAGTCTCCACACAGAACCTGTGAATCGCTCATCACTCAGAACAGATGATGCACA	1689		
QY	1007	CTATGGCTGAGAAAGAGAGCCCTCGAAGAACGAGG-----TCCCTGCCCCATGCC	1061		
Db	1690	GCACCGCCAGTAATTTGTGAAGCTTCAAGTGAATGAGGAGACAGCTTGTATAGATGTC	1749		
QY	1062	TCTTGTCTCCGGAAGTGGATCCCATTTGAAGCCACACCTTTGGCACCAGCACCTCTAC	1121		
Db	1750	ACAAGGATGCTTGGACTGTGCTCACCTGCCTGGTGAAGGACAGGTG--GGTTTCTGGAC	1807		

QY	1122	ACAGTTCTGCATPCTCTTCAAGAGGACCTTCTGTCCATCTCTCAGGACACGGTCTCTGAC	1181
Db	1808	CCAGCTAAGTGTCTGCTTCTGCTGCTCACTGCGCTCCATGTCAGAGATATGTTTGTCTGT	1867
QY	1182	CCACTACGGTTTATGTCCTCCAGTGGTATATGCGTGTCTATCGGCCCTCTCTACTGCA	1241
Db	1868	CCTCATAGGAGTACTACACGGATGTTACTCCCTGGGAGCTACTATCTGTCCAAGT	2107
QY	1242	TATTTGGGACGATGCCAGAGGTTCTCAACAAACACGGCTGCTTCTTCTTCTCTCTCT	1301
Db	1928	GATAGTGGCGATGACAGAAATCTGTCCAATGATATCTGCTGCTTCTTCTTCTGATCT	1987
QY	1302	GTTCCTCATGTTGGCGCCCTCATGCTCACTGCTGCTCACTTCCCTTAGAGATGGCGGT	1361
Db	1988	CTTCGTTGTCGGCGCAATGCCATGCCCTGATCTGCTGTCATGTCAGGACTCGCGGT	2047
QY	1362	TTTCATGAGGAGACCTCAACTACTGTGTACAGCTCAAGCGTATTACCTTGGCCAAAG	1421
Db	2048	GTTCATAGGAGTACTACACGGATGTTACTCCCTGGGAGCTACTATCTGTCCAAGT	2107
QY	1422	CATGCTGACGTGCTTTCAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1481
Db	2108	CCTGCGCATCTGCGCCCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2167
QY	1482	GATGAGGCGCCAGCGCTGAGACAGCGCTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCT	1541
Db	2168	CATGAGGCTGAGCCACCGCAATTCACAGATTTGCCATGCTGCTGCGCCCTGTGGTAT	2227
QY	1542	CACCGCTTGGTGGCCCAATCTTGGGCTGCTGATGTCGAGTGTGCTTCAACTCCTACA	1601
Db	2228	GACCGCTTCAATTTGGCCACTTTCATCGG---GGTATGCGAGGCTCCCTGTTCCACCAT	2284
QY	1602	GGTGGCCACTTTTGGGGCCAGTTACCGCCATCTGCTGCTGCTGCTGCTGCTGCTGCT	1661
Db	2285	GTGGCCATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2344
QY	1662	TGTCAGCTTCAAGACCATCCCACTTACCTGCAATGGAGCTCTTCTCTCTCTCTCTCT	1721
Db	2345	TATTCGGCTGAACAGGCTCTCTGCTTCTGCGCCCATCTCGAGCTGCTCTCTCTCT	2404
QY	1722	GTATGGCTTGTAGGCTGATCTCTGATGATCTATGCTGAGGAGGAGGAGCTGATG	1781
Db	2405	GTACATCTTTCAGGCTGATGCGAGCTCTATGATACGATACGATCGGGGAACTGGAGTG	2464
QY	1782	TTTACAGGAACG-----CTGCGCTTCCGGGAGCCACAGACATCTCCCGAGCGCTGA	1835
Db	2465	CTATGCCACAGCAACTTCTGCTATACAGAACGGGAGGAGTCTCTCAAGGACTTCCA	2524
QY	1836	TGTGGAGATGCCAAGCTTACATGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1895
Db	2525	GATGAGGCAATGATTCGATTTGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCT	2584
QY	1896	GCGGCT 1901	
Db	2585	GCTGCT 2590	

RESULT 12

AAZ98624

ID AAZ98624 standard; cDNA; 2537 BP.

XX

AC AAZ98624;

XX

DT 05-JUN-2000 (first entry)

XX Silkworm Bm white 2 gene.

XX Bm white 2; silkworm; visual characteristic; eye colour; egg colour; ss.

XX Bombyx mori.

XX

PN JP2000000094-A.

XX PD 07-JAN-2000.
XX PF 15-JUN-1998; 98JP-0166944.
XX PR 15-JUN-1998; 98JP-0166944.
XX PA (NORQ) NORINSUISANSHO SANSHI KOMCHU.
XX WPI; 2000-199543/18.
XX P-PSDB; AAY78980.
XX PT Silkorm Bm white gene - for changing visual characters such as egg
XX colour and eye colour of silkworm
XX PS Claim 2; Page 11-13; 16pp; Japanese.
XX CC This sequence represents the silkworm Bm white 2 gene. The Bm white 1, 2,
CC and 3 genes are identified in the invention. Bm white 1 and 2 are
CC homologous to the white gene of humans and mice, and Bm white 3 is
CC homologous to the white gene of the mosquito and fly. The genes can be
CC used for changing visual characteristics such as egg colour and eye
CC colour of the silkworm.

SQ Sequence 2537 BP; 669 A; 646 C; 595 G; 626 T; 1 other;
Query Match 9.4%; Score 324.2; DB 21; Length 2537;
Best Local Similarity 51.7%; Pred No. 2e-70;
Matches 1054; Conservative 0; Mismatches 763; Indels 222; Gaps 5

QY 120 GAAGGTGGAGAACACATCACTGAAGCCACAGCGCTTCTCCACCCTGCCCAAGCGCTCAGC 179
Db 202 GAAGGTGACAAATCACACCGACCCAGCACAGACCCTCAGTCACCTACCGAAAGGACGCC 261
QY 180 CGTGGACATCGATTCTGTGAGCTGTCTTATTCGTCGGGAGGGGCCCTCTCTGGCGCA 239
Db 262 CGTGGATCTGGCATTTACGGACTTGACTTACAAAGTGCAAAAGG-----GAAAGAA 312
QY 240 AAGGGGTATAAGACCCCTTCTCAAGTGCCCTCTCAGGTAAATTCCTGCCCGGGAGCTGAT 299
Db 313 AAGCAATGTCAAACGATCTCAATCTGTTTCGGGTCCGCTCCGCGGAGCTCAC 372
QY 300 TGGCATCATGGGCCCTCAGGGGTGGCAGTCTACATTCATGAACATCTTTGGCAGGATA 359
Db 373 GGCCATCTTGGGTCCGTCGGGTCCGGGAATCAACAATTCACACTACTCACAGGATA 432
QY 360 CAGGGAGCTCGGAATGAAGGGCAGATCTCGTTAATGGAAGGCCACGGGAGCTGAGGAC 419
Db 433 CAACACTTCGGGTATGGAGGGTAGTATAACAGTCAACGGTATGGAGCGTACCTGTCCAG 492
QY 420 CTTCCGCAAGATGTCTGTGTCATCATCATGCAAGATGACATGCTGTCGCGCACCTCAGCGT 479
Db 493 TTTCCGGAAGCTGTCTCTGTCTACATCATGCAAGGATAATCAGCTTCAACGCAACCTTAAC 552
QY 480 GTTGGAGCCATGATGTCTCTGTCTACCTGAATCT-----TACTGAGATCCCGATGT 533
Db 553 CGAAGAGCCATGGCGGTGGCGACATCCCTCAAGCTGCCAAGTTCCACTACGAGAGCGA 612
QY 534 GAAAAAGGATCTCGTGACAGAGATCCTGACGGCACTGGCGGCTGATGCTGCTCCCCACAC 593
Db 613 AAAAGAAATCTGACCGAAACGATCTTTCACACCCCTCAGCTTGATGGAACACCGCTAAGAC 672
QY 594 GAGCAGAGCCCTGCTCTCTGGGGCAGAGAGACCGTCTGGCCATCGCCCTTGGAGCTGGT 653
Db 673 CATGACTTCGAACCTGTCTCGGGGGGTCAAAAGAAACGCTCTCTCCATCGCTCTGGAGTTGGT 732
QY 654 CAACAACCCCGCTGTCTATGTTCTTTGATGAGCCACCAAGTGGTCTGATAGCGGCTCTTG 713
Db 733 CAACAATCCACCTATTATGTCTTCGATGAGCCCTACATCAGGCTGAGACAGTCTTCTCGT 792
QY 714 TTTTCAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGAGGGGGCGGTACCATCATCTGCAC 773
Db 793 TTTTCCAATGATCTCGCTGTCTAAAGAGCGCTCGCAGGAGGGCAGGACCACTCATCTGCAC 852

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:37:55 ; Search time 8632 Seconds
(without alignments)
11648.540 Million cell updates/sec

Title: US-10-072-621-4
Perfect score: 3455
Sequence: 1 gccaccatggcgagagagc.....atgcagttgtattgtggtgg 3455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3426.6	99.2	3498	9	HSA308237	AJ308237 Homo sapi
2	3424.6	99.1	3501	9	HS300465	AJ300465 Homo sapi
3	2761	79.9	2838	9	HS802306	AL137563 Homo sapi
4	2133.4	61.7	3474	10	AF411084	AF411084 Mus muscu
5	2132.2	61.7	3522	10	BC016200	BC016200 Mus muscu
6	2125	61.5	3520	10	MMU426047	AJ426047 Mus muscu
7	1741.4	50.4	157405	2	AP001315	AP001315 Homo sapi
8	1732.8	50.2	1780	9	HS308251	HS308251 Homo sapi
9	1729.4	50.1	219574	9	AP002956	AP002956 Homo sapi
10	1721	49.8	190762	2	AP000833	AP000833 Homo sapi
11	1717.4	49.7	144794	2	AP001182	AP001182 Homo sapi
12	1683.8	48.7	123159	2	HSAC000384	AC000384 Homo sapi
13	1633	47.3	2056	10	AY040865	AY040865 Mus muscu
14	1554.8	45.0	2730	10	BC023077	BC023077 Mus muscu
15	1042.2	30.2	2987	10	MMABC8MR	Z48745 M.musculus
16	1042.2	30.2	3586	6	AX305503	AX305503 Sequence
17	1042.2	30.2	3586	10	MMU34920	U34920 Mus musculus
18	1042.2	30.2	3624	10	AF323659	AF323659 Mus muscu
19	1025.2	29.7	3527	10	RNO303374	AJ303374 Rattus no
20	1015.2	29.4	2996	9	BC029158	BC029158 Homo sapi
21	1012	29.3	2805	9	AF323663	AF323663 Homo sapi
22	1012	29.3	2946	9	AF323662	AF323662 Homo sapi
23	1012	29.3	2983	9	AF323664	AF323664 Homo sapi
24	1012	29.3	3060	9	AF323661	AF323661 Homo sapi
25	1012	29.3	3142	9	AF323660	AF323660 Homo sapi
26	972.8	28.2	2745	9	HSU34919	U34919 Human white
27	972.8	28.2	2930	6	AX331149	AX331149 Sequence
28	972.8	28.2	2930	9	HSWHITE	X91249 H.sapiens m
29	897.8	26.0	3498	9	AY048757	AY048757 Homo sapi
30	747.8	21.6	173657	2	AC112557	AC112557 Rattus no
31	692.8	20.1	249283	2	AC124577	AC124577 Mus muscu
32	669.4	19.4	1327	10	BC026477	BC026477 Mus muscu
33	618.2	17.9	2833	9	AK092959	AK092959 Homo sapi
34	533.2	15.4	687	6	AX430376	AX430376 Sequence
35	429.6	12.4	1497	9	AF038175	AF038175 Homo sapi
36	396.8	11.5	2372	3	AY052113	AY052113 Drosophil
37	324.2	9.4	2537	6	E30646	E30646 Silkworm Bm
38	316.8	9.2	4498	3	AB030450	AB030450 Drosophil
39	313.6	9.1	4184	3	AY119101	AY119101 Drosophil
40	248.4	7.2	2631	3	AY069863	AY069863 Drosophil
41	243.6	7.1	111023	3	AC092242	AC092242 Drosophil
42	243.6	7.1	132666	3	AC005270	AC005270 Drosophil
43	243.6	7.1	218794	2	AC018064	AC018064 Drosophil
44	243.6	7.1	304634	3	AE003576	AE003576 Drosophil
45	243	7.0	3414	3	AY069801	AY069801 Drosophil

ALIGNMENTS

RESULT 1
HSA308237
LOCUS
DEFINITION Homo sapiens mRNA for ABC transporter (ABCG4 gene).
ACCESSION AJ308237
VERSION AJ308237.1 GI:20135821
KEYWORDS ABC transporter; ABCG4 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3498)
AUTHORS Engel, T., Lorkowski, S., Lueken, A., Rust, S., Schlueter, B.,
Berger, G., Cullen, P. and Assmann, G.
TITLE The human ABCG4 gene is regulated by oxysterols and retinoids in

3498 bp mRNA linear PRI 23-APR-2002

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

monocyte-derived macrophages
Biochem. Biophys. Res. Commun. 288, 488-493 (2001)
Engel, T.
Direct Submission
Submitted (25-AUG-2001) Macrophage Metabolism, Institute of
Arteriosclerosis Research, Domagkstrasse 3, Muenster 48149, Germany
Location/Qualifiers
1. .3498

/organism="Homo sapiens"

/db_xref="taxon:9606"

1. .3498

/gene="ABC4"

12. .1952

/gene="ABC4"

/codon_start=1

/product="ABC transporter"

/protein_id="CAC87131.1"

/db_xref="GI:20135822"

/translation="MAEKALEAVGCGLGPVAVAMVLTEDGAEPVLTTHLKKVENHI
TEAQRSHLPKBSADVIEVELSVREGPCPKRKYKTLKCLSKGFCRRELIGIM
PSGAKSTFPMNLAVGREGMKGOILVNGRPRELFRKMSYIMQDDMLPLHVLVE
AMVSNLKLSEKQVKELVTEILFALGMSCHRTALLSGGQRKRLAIALELVNN
PVMFDEPTSLDASCPQVVSLSKSLAQGGRTIITIHQPSAKLFENFDKLYLSQ
GOCIFGVVNLIPYLKGLGLHCPYHNPADFIIEVASGEYDGLNPLFRVQVNGCA
MAEKSSVPEKNEVPAPCPPEVDPIESHTFATSLTOFCILKRTFLSLRDVLT
HLRPMHVVIGVLGLHIGDDASKVENNTGCLFFSMLFLMFAALMETVLTFFLEM
AVFMRHLNYSKAYLAKTMAVDFVQVPCVYCSIVYMTGPAESTRFLFSA
LATATVAOSLGLLIGASNSLOVATFVPTAIPVLLFSGFFVSFRTIPTYLQWS
YLSYVYGEFVILTIYGNRDLTLCLEERCPRFEPOSTLRALDVEDAKLYNDFLVLG
IFFLRLILVLYRVKSR"

BASE COUNT 654 a 1098 c 934 g 812 t

ORIGIN

Query Match 99.2%; Score 3426.6; DB 9; Length 3498;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 3435; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 7 ATGGCGGAGAGCGCTGAGGCGGTGGCTGTGGACTAGGCGGGGGCTGTGGCCATG 66

DB 12 ATGGCGGAGAGCGCTGAGGCGGTGGCTGTGGACTAGGCGGGGGCTGTGGCCATG 71

QY 67 GCCGTGACGCTGAGGACGGGCGGAACCCCTGTGCTGACACGACACCTGAAGAAGTG 126

DB 72 GCCGTGACGCTGAGGACGGGCGGAACCCCTGTGCTGACACGACACCTGAAGAAGTG 131

QY 127 GAGAACCATCAGTGAACCCAGCGCTTCTCCACCTGCCAAGCGCTACGCCGTGGAC 186

DB 132 GAGAACCATCAGTGAACCCAGCGCTTCTCCACCTGCCAAGCGCTACGCCGTGGAC 191

187 ATCGAGTTCGTGAGCTGCTATTCCGTGCGGGGGGGCGCTGTGGCGCAAAAGGGT 246

DB 192 ATCGAGTTCGTGAGCTGCTATTCCGTGCGGGGGGGCGCTGTGGCGCAAAAGGGT 251

QY 247 TATAAGACCTTCTCAAGTGCCTCTCAGGTAATTTCTGCCGCGGGAGCTGATGGCATC 306

DB 252 TATAAGACCTTCTCAAGTGCCTCTCAGGTAATTTCTGCCGCGGGAGCTGATGGCATC 311

QY 307 ATGGGCGCTCAGGGCTGGGAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 366

DB 312 ATGGGCGCTCAGGGCTGGGAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 371

QY 367 TCTGGAATGAAGGGCAGATCTGGTTAATGAAGCCACGAGCTAGGACCTTCGCG 426

DB 372 TCTGGAATGAAGGGCAGATCTGGTTAATGAAGCCACGAGCTAGGACCTTCGCG 431

QY 427 AAGATGTCCTGTACATCATGCAAGATGACATCTGCTGCCACACCTCACGGTGTGGAA 486

DB 432 AAGATGTCCTGTACATCATGCAAGATGACATCTGCTGCCACACCTCACGGTGTGGAA 491

QY 487 GCCATGATGGTCTCTGCTAACCTGAATCTTACTGAGATCCCGATGTGAAAACGATCTC 546

DB 492 GCCATGATGGTCTCTGCTAACCTGAATCTTACTGAGATCCCGATGTGAAAACGATGT 551

QY 547 GTGACAGAGATCTGACGCACTGGCCTGTGCTGCTCCACACGAGGACGACCCCTG 606

DB 552 GTGACAGAGATCTGACGCACTGGCCTGTGCTGCTCCACACGAGGACGACCCCTG 611

QY 607 CTCCTCTGGGGGAGAGAGCGCTCTGGCCATCGCCCTGGAGCTGTGCAACACCGCCCT 666

DB 612 CTCCTCTGGGGGAGAGAGCGCTCTGGCCATCGCCCTGGAGCTGTGCAACACCGCCCT 671

QY 667 GTCATCTTCTTGTATGAGCCACCACTGGTCTGTGATAGCGCTCTTGTTCGAAGTGGT 726

DB 672 GTCATCTTCTTGTATGAGCCACCACTGGTCTGTGATAGCGCTCTTGTTCGAAGTGGT 731

QY 727 TCCCTCATGAAGTCCCTGGCAGAGGGCGGTACCATCATCTGACCACTCCACCAAGCCC 786

DB 732 TCCCTCATGAAGTCCCTGGCAGAGGGCGGTACCATCATCTGACCACTCCACCAAGCCC 791

QY 787 AGTGCCCAAGCTCTTTTGTGATGTTTGTGACAAAGCTCTACATCTGAGCCAGGGTCAATC 846

DB 792 AGTGCCCAAGCTCTTTTGTGATGTTTGTGACAAAGCTCTACATCTGAGCCAGGGTCAATC 851

QY 847 TTCAAGAGAGTGGTCAACCACTGATCCCTTATCTATAAAGGACTCGGCTTGCATTCGCC 906

DB 852 TTCAAGAGTGGTGGTCAACCACTGATCCCTTATCTATAAAGGACTCGGCTTGCATTCGCC 911

QY 907 ACCTTACCAACCCGGCTGACCTTCAATCATCGAGGTGGCTCTGGCGAGTATCGAGACCTG 966

DB 912 ACCTTACCAACCCGGCTGACCTTCAATCATCGAGGTGGCTCTGGCGAGTATCGAGACCTG 971

QY 967 AACCCCATCTTGTTCAGGGCTGTGCAAGTGGGCTGTGCGCTATGGCTGAGAGAAGAGC 1026

DB 972 AACCCCATCTTGTTCAGGGCTGTGCAAGTGGGCTGTGCGCTATGGCTGAGAGAAGAGC 1031

QY 1027 AGCCCTGAGAAGAGAGTCCCTGCCCATGCGCTCTTGTCTCCGGAAGTGGATCCC 1086

DB 1032 AGCCCTGAGAAGAGAGTCCCTGCCCATGCGCTCTTGTCTCCGGAAGTGGATCCC 1091

QY 1087 ATTGAAGCCACACCTTTTCCACACGACCCCTCACACGCTTCTGCACTCTCTTCAAGAG 1146

DB 1092 ATTGAAGCCACACCTTTTCCACACGACCCCTCACACGCTTCTGCACTCTCTTCAAGAG 1151

QY 1147 ACCTTCTCTGCTCATCTCAGGACAGCGTCTGACCCACCTACGGTTCATGTGCCACGTC 1206

DB 1152 ACCTTCTCTGCTCATCTCAGGACAGCGTCTGACCCACCTACGGTTCATGTGCCACGTC 1211

QY 1207 GTTATTGGCGTGTCTATCGGCTCTCTTACCTGCTATATTGGGCGACGATGCCAAGGTC 1266

DB 1212 GTTATTGGCGTGTCTATCGGCTCTCTTACCTGCTATATTGGGCGACGATGCCAAGGTC 1271

QY 1267 TTCAACAACACCGGCTGCTCTTCTTCTCATGCTTCTTCTCATGTCGCGCGCTCATG 1326

DB 1272 TTCAACAACACCGGCTGCTCTTCTTCTCATGCTTCTTCTCATGTCGCGCGCTCATG 1331

QY 1327 CCAACTGTGCTCACCCTTCCCTTAGAGATGGCGGTCTTTCATGAGGAGCACCTCAACTAC 1386

DB 1332 CCAACTGTGCTCACCCTTCCCTTAGAGATGGCGGTCTTTCATGAGGAGCACCTCAACTAC 1391

QY 1387 TGTGTACAGCCTCAAGCGTATTACCTGGCCAGACCATGGCTGACGTCGCCCTTTCAGGTC 1446

DB 1392 TGTGTACAGCCTCAAGCGTATTACCTGGCCAGACCATGGCTGACGTCGCCCTTTCAGGTC 1451

QY 1447 GTGTGTGCGGTGCTTACTGACGATTTGTACTGATGACGCGGCGGCGCTGAGACC 1506

DB 1452 GTGTGTGCGGTGCTTACTGACGATTTGTACTGATGACGCGGCGGCGCTGAGACC 1511

QY 1507 AGCCGCTTCTCTCTTCTCAGCCCTGGCCACCGCACCGCTTGTGTGGCCCAATCTTGT 1566

DB 1512 AGCCGCTTCTCTCTTCTCAGCCCTGGCCACCGCACCGCTTGTGTGGCCCAATCTTGT 1571

QY 1567 GGGCTGTGTATCGGAGCTGTCTTCCAACTCCCTTACAGGTGGGCGACCTTTTGTGGGCGCAGTT 1626

DB 1572 GGGCTGTGTATCGGAGCTGTCTTCCAACTCCCTTACAGGTGGGCGACCTTTTGTGGGCGCAGTT 1631

QY 1627 ACCGCCATCCCTGCTCTTGTTCCTGGCTTCTTTGTCAGCTTCAAGACCATCCCACT 1686
Db 1632 ACCGCCATCCCTGCTCTTGTTCCTGGCTTCTTTGTCAGCTTCAAGACCATCCCACT 1691
QY 1687 TACCTGCAATGAGCTCTATCTCTCTATGTCAGGATGCTTTGAGGCTGTGATCTG 1746
Db 1692 TACCTGCAATGAGCTCTATCTCTCTATGTCAGGATGCTTTGAGGCTGTGATCTG 1751
QY 1747 ACGATCTATGGCATGGAGCGAGGAGACCTTGACATGTTTAGAGGAACGCTGCCCGTTCGG 1806
Db 1752 ACGATCTATGGCATGGAGCGAGGAGACCTTGACATGTTTAGAGGAACGCTGCCCGTTCGG 1811
QY 1807 GAGCACAGAGCATCTCCGAGCGCTGATGATGAGAGTGCACAGCTCTACATGACATTC 1866
Db 1812 GAGCACAGAGCATCTCCGAGCGCTGATGATGAGAGTGCACAGCTCTACATGACATTC 1871
QY 1867 CTGCTGTTGGGATCTCTCTCTAGCCCTGGGCTGCTGGCTACCTTGTGCTGCTGCTAC 1926
Db 1872 CTGCTGTTGGGATCTCTCTCTAGCCCTGGGCTGCTGGCTACCTTGTGCTGCTGCTAC 1931
QY 1927 CGGTCAGATCAGAGATAGAGGCTTGCCCGAGGCTGTACCCAGCCCTGTCAGCAGGA 1986
Db 1932 CGGTCAGATCAGAGATAGAGGCTTGCCCGAGGCTGTACCCAGCCCTGTCAGCAGGA 1991
QY 1987 AGCCCCAGTCCAGCCCTTTGGGACTGTTTAACTTATAGACTTGGGCTGCTGCTCT 2046
Db 1992 AGCCCCAGTCCAGCCCTTTGGGACTGTTTAACTTATAGACTTGGGCTGCTGCTCT 2051
QY 2047 GCGGGGCTATCT 2106
Db 2052 GCGGGGCTATCT 2111
QY 2107 AGCTGGGCTCTGGAGTGGGGCTTCCAGCCCTCCCACTATGCCAGGAGTCTTCCCAA 2166
Db 2112 AGCTGGGCTCTGGAGTGGGGCTTCCAGCCCTCCCACTATGCCAGGAGTCTTCCCAA 2171
QY 2167 GTTGATGGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2226
Db 2172 GTTGATGGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2231
QY 2227 GGAGCTCTGCT 2286
Db 2232 GGAGCTCTGCT 2291
QY 2287 AGGCTCTCTAGGGCCCCACTTACAACTGACCAAGTGGGCCCTCTGGGGTCCCAACA 2346
Db 2292 AGGCTCTCTAGGGCCCCACTTACAACTGACCAAGTGGGCCCTCTGGGGTCCCAACA 2351
QY 2347 CACAAGTGTGTAAGTGGGCTGCTATAAGTGGGCTGCTATAAGTGGGCTGCTATAAG 2406
Db 2352 CACAAGTGTGTAAGTGGGCTGCTATAAGTGGGCTGCTATAAGTGGGCTGCTATAAG 2411
QY 2407 AGTCCACTGGAGTCCCATATGATGTTGAATGACAGGGAAGGACTCTGGAAGTCTC 2466
Db 2412 AGTCCACTGGAGTCCCATATGATGTTGAATGACAGGGAAGGACTCTGGAAGTCTC 2471
QY 2467 TTCT 2526
Db 2472 TTCT 2531
QY 2527 ACAGAGCTGGGTTTCTGCTAGTGCACCTCTCCCAATCTCTGGGGATGAGAGGCTG 2586
Db 2532 ACAGAGCTGGGTTTCTGCTAGTGCACCTCTCCCAATCTCTGGGGATGAGAGGCTG 2591
QY 2587 GGGCTGTGGGATGCCCATCCCTCTCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2646
Db 2592 GGGCTGTGGGATGCCCATCCCTCTCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2651
QY 2647 ACCTGTGCAATAATCTGTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2706
Db 2652 ACCTGTGCAATAATCTGTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2711
QY 2707 TATCTGGGCGGGGCTGAGTGGGGAAGACCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 2766

Db 2712 TATCTGGGCGGGGCTGAGTGGGGAAGACCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2771
QY 2767 CATGCACGCTCTGCTGCT 2826
Db 2772 CATGCACGCTCTGCTGCT 2831
QY 2827 GGCATCTCAGCCTTACACACTTGGCATCCCTTACAGCAGAGAGAGAGTGTGATGGCTG 2886
Db 2832 GGCATCTCAGCCTTACACACTTGGCATCCCTTACAGCAGAGAGAGAGTGTGATGGCTG 2891
QY 2887 ATGCTGTGGTGGGCTGCTGCTGGGAGGACAGTGCACCACTCTCTCTCTCTCTCTCTCTCT 2946
Db 2892 ATGCTGTGGTGGGCTGCTGCTGGGAGGACAGTGCACCACTCTCTCTCTCTCTCTCTCTCT 2951
QY 2947 GTTGAGACTCTAAGGATAAGGCTGCTGCTGCCAGGCTGTCTACAGGAACCTGCAGGCTG 3006
Db 2952 GTTGAGACTCTAAGGATAAGGCTGCTGCTGCCAGGCTGTCTACAGGAACCTGCAGGCTG 3011
QY 3007 CTACCCCCAAGCTCTTCCCTCTCCCAAGCAGGGGTGGCAGAGGCTAGATCCCTGGA 3066
Db 3012 CTACCCCCAAGCTCTTCCCTCTCCCAAGCAGGGGTGGCAGAGGCTAGATCCCTGGA 3071
QY 3067 GTTCAGGAACCAACAGCAGCAGCAGGCGATAGTTGGCTTGGGCTGAGGCTGAGGCTG 3126
Db 3072 GTTCAGGAACCAACAGCAGCAGCAGGCGATAGTTGGCTTGGGCTGAGGCTGAGGCTG 3131
QY 3127 CGGCT 3186
Db 3132 CGGCT 3191
QY 3187 GCTGCTCATTTAACTTCTGCT 3246
Db 3192 GCTGCTCATTTAACTTCTGCT 3251
QY 3247 GCATCAGAACAGCTGCT 3306
Db 3252 GCATCAGAACAGCTGCT 3311
QY 3307 TTCTCTGGTGGGCTCAGGCTCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3366
Db 3312 TTCTCTGGTGGGCTCAGGCTCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3371
QY 3367 CCCTTTTCTAT 3426
Db 3372 CCCTTTTCTAT 3431
QY 3427 CACACAGGATGCCAGTGTGATTTGTTGGG 3455
Db 3432 CACACAGGATGCCAGTGTGATTTGTTGGG 3460

RESULT 2
HSA300465
LOCUS
DEFINITION
Homo sapiens mRNA for putative white family ATP-binding cassette transporter (ABCG4 gene).
ACCESSION
AJ300465
VERSION
AJ300465.1
KEYWORDS
ABCG4 gene; white family ATP-binding cassette transporter.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Oldfield, S., Lowry, C.A. and Lightman, S.L.
TITLE
ABCG4: a novel white family ATP-binding cassette transporter expressed in human hypothalamus
JOURNAL
Abstr. - Soc. Neurosci. 27, 1951-1951 (2001)
REFERENCE
2 (bases 1 to 3501)
AUTHORS
Oldfield, S.
TITLE
Direct Submission
JOURNAL
Submitted (22-NOV-2000) Oldfield S., University Research Centre for

Neuroendocrinology, University of Bristol, Bristol royal Infirmary,
Marlborough Street, Bristol, BS2 8HW, UNITED KINGDOM

FEATURES
Source

1..3501

/organism="Homo sapiens"

/db_xref="RZPD:IMAGp998E133893"

/db_xref="taxon:9606"

/clone="I.M.A.G.E. consortium clone ID 1537140"

/tissue_type="Dorsal root ganglion"

1..3501

/gene="ABCG4 gene"

1..56

/gene="ABCG4 gene"

57..1940

/gene="ABCG4 gene"

/codon_start=1

/product="putative white family ATP-binding cassette

transporter"

/protein_id="CAC17140.1"

/db_xref="GI:11342541"

/db_xref="SWISS-PROT:O9H172"

/translation="NAVTLDEGAEPVLTLLKLVENHTEAQRSHLPKRSADVIEP

VELSVREGPCWRKRGKTLKLSGKFCRELIGIMPSPGAGKSTFMNIIAGYRES

GKGGIIVNGRPRLTRFRKMSCYIMQDDMLPHLTVLEAMVSNLKLSEKQYKKE

LVTETLALGLMSCSHRTALLSGGQRRLALELVNPNVPMFDEPTSGIDSASF

QVSLMKSLAOGGRTICTIHQPSAKLFEMFKLYILSGOCIFKGVVNLIPYKGL

GLHCPTYHNPADFIIEVASGEYGLNPLMFRAVONGLCAMAEKSSPEKNEVPAPCP

CPPEVDPIESHTFATSTQPCILEFKRTFLSLRDLVTLHLPMSHVIVGLIGLLYL

HIGDDAKFVNNTGCLFMSLMFLMFAALPVLTFPLEMAFMFRLNHWLSKAYLL

AKTMDADVQVCPVVCISIVVMYTGQPAETSRFLFSALATATLVAQSLGLLIGAA

SNLQVATFEGPVTAPVLLVLSGFVSKFTIPTYLQWSSYLSVRYGFEVILTIYGM

ERGDITCLEERCFFPEPOSILLRALDVEDAKLYMDFLVLGIFFLALRLLAYLVLRVK

SER"

1941..3501

/gene="ABCG4 gene"

668 a 1095 c 927 g 810 t 1 others

3'UTR
BASE COUNT
ORIGIN

Query Match 99.1%; Score 3424.6; DB 9; Length 3501;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 3433; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 8 TGGCGGAGAGCGCTGGAGGCGGTGGGCTGTGGACTAGGGCCGGGGCGGTGGCCATGG 67
DB 1 TGGCGGAGAGCGCTGGAGGCGGTGGGCTGTGGACTAGGGCCGGGGCGGTGGCCATGG 60

QY 68 CCGTGACGCTGGAGGACGGGGCGGAACCCCTGTCTGCTGACACGACCTGAAGAGGTGG 127
DB 61 CCGTGACGCTGGAGGACGGGGCGGAACCCCTGTCTGCTGACACGACCTGAAGAGGTGG 120

128 AGAACCATCATCTAAGCCGAGCGCTTCTCCACCTGCGCCAGCGCTCAGCGCTGGACA 187
121 AGAACCATCATCTAAGCCGAGCGCTTCTCCACCTAAGCGCTCAGCGCTGGACA 180

QY 188 TGGAGTTCTGGAGCTGCTCTAATCCGTGCGGGAGGGGCGCTGTGGCGCAAAAGGGTT 247
DB 181 TGGAGTTCTGGAGCTGCTCTAATCCGTGCGGGAGGGGCGCTGTGGCGCAAAAGGGTT 240

QY 248 ATAAGACCTTCTCAAGTGCCTCTCAGCTAATCTGCGCGGAGGCTGATGCGATCA 307
DB 241 ATAAGACCTTCTCAAGTGCCTCTCAGCTAATCTGCGCGGAGGCTGATGCGATCA 300

QY 308 TGGGCGCTCAGGGCTGGAGCTACATTCATGAACATCTTGGCAGATACAGGGAGT 367
DB 301 TGGGCGCTCAGGGCTGGAGCTACATTCATGAACATCTTGGCAGATACAGGGAGT 360

QY 368 CTGGAATGAAGGGGAGAGCTCTGGTTAATGAAGGCCACGGGAGCTGAGGACCTTCCGCA 427
DB 361 CTGGAATGAAGGGGAGAGCTCTGGTTAATGAAGGCCACGGGAGCTGAGGACCTTCCGCA 420

QY 428 AGATGCTCTGCTACATCATGAAGATGACATGCTCTGCTGCGGACCTCAGGTTGGAG 487
DB 421 AGATGCTCTGCTACATCATGAAGATGACATGCTGCTGCGGACCTCAGGTTGGAG 480

QY 488 CCATGATGCTCTGCTAACTGAATCTTACTGAGATCCCGATGTGAAAAAGCATCTCG 547
DB 481 CCATGATGCTCTGCTAACTGAAGCTGAGTGAGAGCAGGAGGAGGAGCTGG 540

QY 548 TGACAGAGATCTGACGGCACTGGGCTGATGCTGCTGCTCCACACAGGAGGAGCGCTCG 607
DB 541 TGACAGAGATCTGACGGCACTGGGCTGATGCTGCTGCTCCACACAGGAGGAGCGCTCG 600

QY 608 TCTCTGGGGGAGAGGAGGCTGCGGCATCGGCCTGGAGTGCTGATGCTGCTGCTGCTGCTG 667
DB 601 TCTCTGGGGGAGAGGAGGCTGCGGCATCGGCCTGGAGTGCTGATGCTGCTGCTGCTGCTG 660

QY 668 TCATGTTCTTTCATGAGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
DB 661 TCATGTTCTTTCATGAGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

QY 728 CCTCATGAAGTCCCTGGCACAGGGGCGCTGACCATCATCTGACCATCATCTGACCATCAT 787
DB 721 CCTCATGAAGTCCCTGGCACAGGGGCGCTGACCATCATCTGACCATCATCTGACCATCAT 780

QY 788 GTGCCAAGCTCTTTCATGATGTTTGACAGCTCTACATCTGACGAGGCTGATGCTGATCT 847
DB 781 GTGCCAAGCTCTTTCATGATGTTTGACAGCTCTACATCTGACGAGGCTGATGCTGATCT 840

QY 848 TCAAAAGGAGTGTTCACCACTGATCCCTATCTAAAGGAGCTCGGCTTGCATTTGCCCA 907
DB 841 TCAAAAGGAGTGTTCACCACTGATCCCTATCTAAAGGAGCTCGGCTTGCATTTGCCCA 900

QY 908 CCTACCAACACCGGCTGACTTCATCATGAGGAGGCGCTCTGCGAGTATGAGAGCTGA 967
DB 901 CCTACCAACACCGGCTGACTTCATCATGAGGAGGCGCTCTGCGAGTATGAGAGCTGA 960

QY 968 ACCCATGTTGTTGAGGCTGTGACAGATGGGCTGTGGCTGTGGCTGTGAGAGAGAGCA 1027
DB 961 ACCCATGTTGTTGAGGCTGTGACAGATGGGCTGTGGCTGTGGCTGTGAGAGAGAGCA 1020

QY 1028 GCGCTGAGAGAAACGAGGTCCCTGCGCCATGCGCTTGTCTTCTCCGGAAGTGGATCCCA 1087
DB 1021 GCGCTGAGAGAAACGAGGTCCCTGCGCCATGCGCTTGTCTTCTCCGGAAGTGGATCCCA 1080

QY 1088 TTGAAAGCCACACCTTTGCCACAGCAGCCCTCACACAGTTCATCTCTTCAAGAGGA 1147
DB 1081 TTGAAAGCCACACCTTTGCCACAGCAGCCCTCACACAGTTCATCTCTTCAAGAGGA 1140

QY 1148 CTTTCTCTGTCATCTCAGGACAGGCTCTGACCCACCTAGGTTTCATGCTCCAGGTGG 1207
DB 1141 CTTTCTCTGTCATCTCAGGACAGGCTCTGACCCACCTAGGTTTCATGCTCCAGGTGG 1200

QY 1208 TTATTTGGGCTGCTCATCGGCTTCTTACCTGACATATTGGCGACGATGCGCAGGCTCT 1267
DB 1201 TTATTTGGGCTGCTCATCGGCTTCTTACCTGACATATTGGCGACGATGCGCAGGCTCT 1260

QY 1268 TCAACAAACACGGCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1327
DB 1261 TCAACAAACACGGCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1320

QY 1328 CAACCTGCTCAGCTTCCCTTAGAGATGGCGGCTTCTCATGAGGAGGACCTCAACTACT 1387
DB 1321 CAACCTGCTCAGCTTCCCTTAGAGATGGCGGCTTCTCATGAGGAGGACCTCAACTACT 1380

QY 1388 GGTACAGCTCAAAGCGTATTACCTGGCGCAAGACCATGCTGACCTGCGCTTTCAGGTGG 1447
DB 1381 GGTACAGCTCAAAGCGTATTACCTGGCGCAAGACCATGCTGACCTGCGCTTTCAGGTGG 1440

QY 1448 TGTGTCGGGTGCTTACTGACAGATTTGTACTTGGATGACGGGCGAGCCGCTGAGACCA 1507
DB 1441 TGTGTCGGGTGCTTACTGACAGATTTGTACTTGGATGACGGGCGAGCCGCTGAGACCA 1500

QY 1508 GCGGCTTCTCTCTCTCAGCCCTGGCCACCGGCTTGGTGGCCCAATCTTTGG 1567
DB 1501 GCGGCTTCTCTCTCTCAGCCCTGGCCACCGGCTTGGTGGCCCAATCTTTGG 1560

QY 1568 GGCTGCTGATCGAGAGTCTTCCAACTCCCTACAGGTGGCCACTTTTGTGGGCCAGTTA 1627
Db 1561 GGTCTGATCGGAGTCTTCCAACTCCCTACAGGTGGCCACTTTTGTGGGCCAGTTA 1620
QY 1628 CCGCCATCCCTGCTCTTCTTCTCGGCTTCTTGTGACGTTCAAGACCATCCCACTT 1687
Db 1621 CCGCCATCCCTGCTCTTCTTCTCGGCTTCTTGTGACGTTCAAGACCATCCCACTT 1680
QY 1688 ACCTGCAATGGAGTCCCTATCTCTCTATGTCAGGTATGGCTTTGAGGGTGTGATCTGA 1747
Db 1681 ACCTGCAATGGAGTCCCTATCTCTCTATGTCAGGTATGGCTTTGAGGGTGTGATCTGA 1740
QY 1748 CGATCTATGCTAGGAGGAGAGACCTGACATGTTTAGAGAACGCTGCCGTTCCGGG 1807
Db 1741 CGATCTATGCTAGGAGGAGAGACCTGACATGTTTAGAGAACGCTGCCGTTCCGGG 1800
QY 1808 AGCCACAGACATCTCCCGAGCCTGGATGTGAGGATGCCAAGCTCTACATGACTTC 1867
Db 1801 AGCCACAGACATCTCCCGAGCCTGGATGTGAGGATGCCAAGCTCTACATGACTTC 1860
QY 1868 TGGTCTTGGGCACTTCTTCTTCTAGCCCTGGGCTGCTGGCTTACCTTGTCTGCTTTACC 1927
Db 1861 TGGTCTTGGGCACTTCTTCTTCTAGCCCTGGGCTGCTGGCTTACCTTGTCTGCTTTACC 1920
QY 1928 GGGTCAAGTCAGAGAGATAGAGGTTGCCCCAGCCTGTACCCAGCCCTGACAGAGAA 1987
Db 1921 GGGTCAAGTCAGAGAGATAGAGGTTGCCCCAGCCTGTACCCAGCCCTGACAGAGAA 1980
QY 1988 GCCCCAGTCCAGCCCTTTGGGACTGTTTAACTTTATAGACTTGGGCACTGTTCTCG 2047
Db 1981 GCCCCAGTCCAGCCCTTTGGGACTGTTTAACTTTATAGACTTGGGCACTGTTCTCG 2040
QY 2048 GCGGGGCTATCTCTCTCTCTTGGCTCTCCACAGGCTGGCTGCGGACTGGGCTCCCA 2107
Db 2041 GCGGGGCTATCTCTCTCTCTTGGCTCTCCACAGGCTGGCTGCGGCTCCCA 2100
QY 2108 GCCTGGGCTCTGGAGTGGGGCTCCAGCCCTGCCCACTATGCCAGGAGTCTTCCCAAG 2167
Db 2101 GCCTGGGCTCTGGAGTGGGGCTCCAGCCCTGCCCACTATGCCAGGAGTCTTCCCAAG 2160
QY 2168 TTGATCGGTTTGTAGCTTCTCTCTCTCTCTCAACACTGTCATGCAAGACTACTGG 2227
Db 2161 TTGATCGGTTTGTAGCTTCTCTCTCTCTCTCTCAACACTGTCATGCAAGACTACTGG 2220
QY 2228 GAGGCTGCTGCTCTCTCTCTGCCCCATGGACCCCTCTCTCTCTGCTGCTGCTGCTGCTG 2287
Db 2221 GAGGCTGCTGCTCTCTCTCTGCCCCATGGACCCCTCTCTCTCTGCTGCTGCTGCTGCTG 2280
QY 2288 GGCTCTCTAGGCCCCACTTACAACTGACAAAGTGCCGCCCTCTCTGGGGTCCCCACCAC 2347
Db 2281 AGCTCTCTAGGCCCCACTTACAACTGACAAAGTGCCGCCCTCTCTGGGGTCCCCACCAC 2340
QY 2348 ACAAGTCTTGTAACTGGGCTGCTATAAGTGTGGAGTTCACAGGCTGGGCCCTGGTGA 2407
Db 2341 ACAAGTCTTGTAACTGGGCTGCTATAAGTGTGGAGTTCACAGGCTGGGCCCTGGTGA 2400
QY 2408 GTCCACTGGAAGTCCCAATTATGATGTTGAAATGGACAGGAGGACTCTGGAAGTCTCT 2467
Db 2401 GTCCACTGGAAGTCCCAATTATGATGTTGAAATGGACAGGAGGACTCTGGAAGTCTCT 2460
QY 2468 TCCT 2527
Db 2461 TCCT 2520
QY 2528 CAGAGCTGGGTTTCTGCTAGTCTACCTACCTCCCAATCTGGGATGGAGGCTGG 2587
Db 2521 CAGAGCTGGGTTTCTGCTAGTCTACCTACCTCCCAATCTGGGATGGAGGCTGG 2580
QY 2588 GGCTGGGATGCCCAATCCCT 2647
Db 2581 GGCTGGGATGCCCAATCCCT 2640
QY 2648 CCGTGTGCAATATGCTGCTGTTCTCTCCCACTGGCACTGGAACTGGAAATGCACATTT 2707

Db 2641 CCTGTGCAATATGCTGCTGTTCTCTCTCCCACTGGCACTGGAAATGCACATTT 2700
QY 2708 ATTCTGGGGGGGTGAGTGGGGAAGACCCAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2767
Db 2701 ATTCTGGGGGGGTGAGTGGGGAAGACCCAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
QY 2768 ATCAGGGTCTCGTATGCT 2827
Db 2761 ATCAGGGTCTCGTATGCT 2820
QY 2828 GCCATCTCAGCCCTACACACTTCCCATCTCCCTCTACAGCACAGAGAGATGATGGTGGCA 2887
Db 2821 GCCATCTCAGCCCTACACACTTCCCATCTCCCTCTACAGCACAGAGAGATGATGGTGGCA 2880
QY 2888 TGTCTGTGTGGGGGTGCTGGTGGGAGGACAGTCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2947
Db 2881 TGTCTGTGTGGGGGTGCTGGTGGGAGGACAGTCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
QY 2948 TTGGAGACTCTAAGGATAAGGCTGGTCTGCCAGGCTGTCTACAGAACTGCAGGTGTC 3007
Db 2941 TTGGAGACTCTAAGGATAAGGCTGGTCTGCCAGGCTGTCTACAGAACTGCAGGTGTC 3000
QY 3008 TACCCCAAGTCTTCCCTCTCCCAAGCCAGGGTGGCACAGGCACTAGATCCTCTGGAG 3067
Db 3001 TACCCCAAGTCTTCCCTCTCCCAAGCCAGGGTGGCACAGGCACTAGATCCTCTGGAG 3060
QY 3068 TTCAAGAACCAACACAGACCAACACCGGCACTAAGTTGGCTTGGCCACTGCCACCCAC 3127
Db 3061 TTCAAGAACCAACACAGACCAACACCGGCACTAAGTTGGCTTGGCCACTGCCACCCAC 3120
QY 3128 GGCCCTCTCTTGTGTCT 3187
Db 3121 GGCCCTCTCTTGTGTCT 3180
QY 3188 CTGCTCATCAAACTTCTGTCT 3247
Db 3181 CTGCTCATCAAACTTCTGTCT 3240
QY 3248 GCATCAGAACAGCTGCTGCTGGCCAGGCTGGCACACACTCAGAGCATGTCTGGCTT 3307
Db 3241 GCATCAGAACAGCTGCTGCTGGCCAGGCTGGCACACACTCAGAGCATGTCTGGCTT 3300
QY 3308 TCCTGTGGTCCAGGCTCATCT 3367
Db 3301 TCCTGTGGTCCAGGCTCATCT 3360
QY 3368 CCTTTTCT 3427
Db 3361 CCTTTTCT 3420
QY 3428 ACACAGGATGCCAGTTGATTTTGGG 3455.
Db 3421 ACACAGGATGCCAGTTGATTTTGGG 3448

RESULT 3

HSR802306 2838 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKF2p434p1420 (from clone DKF2p434p1420);
DEFINITION partial cds.
ACCESSION AL137563
VERSION AL137563.1 GI:6808269
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and
Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152

COMMENT	Martinsried, GERMANY		QY	1115	CCCTCACACAGTTCTGCATCTCTTCAAGAGGACCTTCTCTGTCCATCTCCTCAGGACACAGG	1174
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by Medicinix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp434P1420) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.			Db	421	CCCTCACACAGTTCTGCATCTCTTCAAGAGGACCTTCTGTCCATCTCCTCAGGACACAGG
FEATURES	Location/Qualifiers		QY	1175	TCCTGACCCACCTACAGTTTCATGTCGCCACGTTGTTATGGGGTGTCTCATCGGCTCTCTCT	1234
	1. .2838		Db	481	TCCTGACCCACCTACAGTTTCATGTCGCCACGTTGTTATTTGGCGTGTCTCATCGGCTCTCTCT	540
source	/organism="Homo sapiens"		QY	1235	ACCTGCATATTTGGCGACCATGCCAGCAAGTCTTCAACACACCGGCTGCTCTCTCTCTCT	1294
	/db_xref="taxon:9606"		Db	541	ACCTGCATATTTGGCGACCATGCCAGCAAGTCTTCAACACACCGGCTGCTCTCTCTCTCT	600
gene	/map="11"		QY	1295	CCATGCTGTTTCCATGTTTCGCGCGCTCATGTCGCCAACTGTGCTCACCTTCCCTTAGAGA	1354
	/clone="DKFZp434P1420"		Db	601	CCATGCTGTTTCCATGTTTCGCGCGCTCATGTCGCCAACTGTGCTCACCTTCCCTTAGAGA	660
CDS	/tissue_type="testis"		QY	1355	TGGCGTCTTCATGAGGAGGACCTCAACTACTGGTACAGCCTCAAAAGCGTATTACTCTGG	1414
	/clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"		Db	661	TGGCGTCTTCATGAGGAGGACCTCAACTACTGGTACAGCCTCAAAAGCGTATTACTCTGG	720
	/dev_stage="adult"		QY	1415	CAAAGACCATGGCTGACGTGCGCTTTCAGTGTGTTGTCGGTGGTCTACTGAGCATTTG	1474
	1. .1253		Db	721	CAAAGACCATGGCTGACGTGCGCTTTCAGTGTGTTGTCGGTGGTCTACTGAGCATTTG	780
	/gene="DKFZp434P1420"		QY	1475	TGCTACTGGATGACGGCGCGCTGAGACACGCGCTTCTGCTCTCTCTCAGCCCTGG	1534
	<1. .1253		Db	781	TGCTACTGGATGACGGCGCGCTGAGACACGCGCTTCTGCTCTCTCTCAGCCCTGG	840
polyA_signal	/note="strong similarity to mouse white protein"		QY	1535	CCACGCCACCGCTTGTGTGCCCAATCTTTGGGGTGTCTGATCGGAGCTGCTTCCAAC	1594
	/codon_start=3		Db	841	CCACGCCACCGCTTGTGTGCCCAATCTTTGGGGTGTCTGATCGGAGCTGCTTCCAAC	900
polyA_site	/product="hypothetical protein"		QY	1595	CCCTACAGTGGCCACTTTTGTGGGCCAGTTACGGCCATCCCTGTCTCTTGTCTCCG	1654
	/protein_id="CA870814.1"		Db	901	CCCTACAGTGGCCACTTTTGTGGGCCAGTTACGGCCATCCCTGTCTCTTGTCTCCG	960
ORIGIN	/db_xref="GI:6808270"		QY	1655	GCTTCTTGTGACGTTCACAGACCATCCCACTTACCTGCAATGGAGCTCTCTATCTCTCT	1714
	/translation="LDSASCFQVVSIMKSLAQGRITICTIHOPSAKLFEFMDKLVIL SQQCIFKGVNLPYLLGLHCPTNHPADPIIEVASGEYGDINPMLFRAVONG L					

[illegible]

Db	2581	AGGTGGGAGACACACTCAGACGATGCTGGCTTTTCCGTGGGTCCAGGCTCATTCCTGCT	2640
QY	3335	TCTGATTTCCCTCCCGCCAGGGCTCATTTTCCCGCTTTTCCCTGCTACACATCCCTGCTCA	3394
Db	2641	TCTGATTTCCCTCCCGCCAGGGCTCATTTTCCCGCTTTTCCCTGCTACACATCCCTGCTCA	2700
QY	3395	CCTCCTCTCACCCCTGCCACAGATCTTCTCTATCATCACAGGATGCCAGTTGTATTTCTGG	3454
Db	2701	CCTCCTCTCACCCCTGCCACAGATCTTCTCTATCATCACAGGATGCCAGTTGTATTTCTGG	2760
QY	3455	G	3455
Db	2761	G	2761
RESULT 4			
LOCUS	AF411084		
DEFINITION	Mus musculus ATP-binding cassette transporter sub-family G member 4 (Abcg4) mRNA, complete cds.	3474 bp	mRNA linear ROD 27-MAR-2002
ACCESSION	AF411084		
VERSION	AF411084.1	GI:18034136	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE			
AUTHORS	1 (bases 1 to 3474)		
TITLE	Human and mouse orthologs of a new ATP-binding cassette gene, ABCG4		
JOURNAL	Cytogenet. Cell Genet. 94 (3-4), 196-201 (2001)		
MEDLINE	21845887		
PUBMED	11856881		
REFERENCE	2 (bases 1 to 3474)		
AUTHORS	Annilo, T., Tammaru, J., Hutchinson, A., Rzhetsky, A., Dean, M. and Allikmets, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-AUG-2001) Laboratory of Genomic Diversity, National Cancer Institute, NCI-Frederick, Frederick, MD 21702, USA		
FEATURES			
source	Location/Qualifiers		
gene	1..3474		
CDS	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	1..3474		
	/gene="Abcg4"		
	122..2062		
	/gene="Abcg4"		
	/note="ABCG4"		
	/codon_start=1		
	/product="ATP-binding cassette transporter sub-family G member 4"		
	/protein_id="AA157369.1"		
	/db_xref="GI:18034137"		
	/translation="MAEKALAEAVCGGLGPGAVAMAVTLEDGAEPVLVTLHLKKVENHIT TEAFRSHLPKRSADVTEFVELSYSVREGPCWRKRGYKTLKCLSGKFCRELIGMGI PSGAGKSTFNILAGYRESGKMGQILYNGRPRELTERKMSCYIMQDMLPLHLTVLE AMWVSANLKSEKQEVKRELVEITLALGMSKRTIALSGGQKRLAIALELVNN PPVMEFPTSGDLSACFQVYSLMSLAHGGRTVICTIHQPSAKTFEMFDKLYLGC QCQIFKGVNLIPIYLRKGLHCHPTYNPADFIIEVASGEYGLNPLMFRAVQNGLC MAEKKSSPGNELPAHCTGCPDQIESHTFATSTLQFCILFRRTFLTLRDVTLT HLRMSHVLIGVLTGLLYLHGDDASKVENNTGFLFFSMFLFAALMPTVLPFLPELM AVFREHLNHWYTLKAYILAKTMADVFQVYVCPVYCSIVYVMTGQPAETSRFLLFSAS LATATVAQSLGULLIGAASVLSQVATFVGPTVAIPVLLFSGFVSFKTIPTYLQWSS YLSYRVYFEGELIITYGMERGHITCLDEQCPRDPQIILRELDVEEAKLYMDFLVILG IFFLALRLALLVILRVYK58R"		
BASE COUNT	648 a	1073 c	915 g 838 t
ORIGIN			
Query Match	61.7%	Score 2133.4;	DB 10; Length 3474;
Best Local Similarity	80.4%;	Prod. No. 0;	
Matches 2775;	Conservative	0; Mismatches 566;	Indels 112; Gaps 19;

Qy	7	ATGCGGAGAAAGCGCTTGGAGCCCTGGGCTGTGGCTGTGGACTAGGGCCGGGGGCTGTGGCCATG	66
Db	122	ATGCGGAGAAAGCAATTGAGGCGCTGGGCTGTGGACTGGGCCCTGGGCTGTGGCCATG	181
Qy	67	GCCCTGACGCTTGGAGGAGCGGGCGGAACCCCTGCTGTGTGACCGACCGACCTGAAGAAGGTG	126
Db	182	GCTGTGACACTTGGAGAGCGGGGAGAGCCCCCTGTGCTGTGACCGACCGACCTAAAGAAGGTG	241
Qy	127	GAGAACACATCACTGAAGCCACCGCTTCTCCACCTTGCCCAAGCGCTCAGCCGTGGAC	186
Db	242	GAGAACCATATCACAGAAGCACACCGCTTCTCTCACTGCCAAGCGCTCTGCTGTGGAC	301
Qy	187	ATCAGATTCTGTGGAGCTGTCTTATTCCTGTGGGAGGGGCCCTGCTGGCGCAAAAGGGT	246
Db	302	ATCAGATTCTGTGGAACTGTCTTATTCCTGTGGGAGGGGCCCTGCTGGCGCAAAAGGGT	361
Qy	247	TATAAGACCCTTCTCAAGTGCCTCTCAGGTAAATTTCTGCCCGGGAGCTGATTGGCATC	306
Db	362	TACAAGACCCTTCTCAAGTGCCTCTCTGTTAAGTTTCTGTGCCGAGAGCTGATTGGCATC	421
Qy	307	ATGGGCCCTCAGGGCTGGCAAGTCTACATTTCATGAACATCTTTGGCAGGATACAGGGAG	366
Db	422	ATGGGCCCTCTGGGGCTGGCAAGTCCACATTCTATGAACATCTTTGGCAGGTTACAGGGAG	481
Qy	367	TCTGGAAATGAAGGGGAGATCCTGGTTAATGGAAGCCACGGGAGCTGAGGACCTTCCGC	426
Db	482	TCGGGGATGAAGGGGAGATCCTGGTTAAACGGGAGGCGGCGAGCTGAGGACCTTCCGC	541
Qy	427	AAGATGTCTGTCTACATCATGCAAGATGACATGCTGCTGGCGCACTTCACGGTCTTGGAA	486
Db	542	AAGATGTCTGTCTACATCATGCAAGATGACATGCTGCTGCCACCTTCACAGTCTCTGGAG	601
Qy	487	GCCATGATGTCTCTGCTAACTGAAATCTTACTGAGAATCCCGATGTGAAAACAGTCTC	546
Db	602	GCCATGATGTCTCTGCTAACTGAAATCTGAGTGAAGAAGAGGTGAAGAAGAACTG	661
Qy	547	GTGACAGAGATCCTGACGGCACTGGGCGCTGATGTCTGTCTCCACACGAGGACAGCCCTG	606
Db	662	GTGACAGAGATCCTGACAGCCCTGGGACTGATGTCTGTCTCTCACAGGAGACAGCTCTG	721
Qy	607	CTCTCTGGCGGGCAGAGAAAGCGTCTGGCCATCGCCCTGGAGCTGTGTCAACAACCCGCT	666
Db	722	CTCTCTGGAGGCGAGAGAGCGCTTGCCATTTGCTCTGGAAGTGGTCAACAACCCGCT	781
Qy	667	GTCATGTCTTTGATGAGCCACACAGTGTGTGGATAGCGCCCTGTTTCCAAAGTGGTG	726
Db	782	GTCATGTCTTTGATGAGCCATACCAGCGGTCTAGACAGCGGTCTTCTGTTCCAAAGTGGTG	841
Qy	727	TCCTCATGAAGTCCCTGGCACAGGGGGCGGTACCATCATCTGCACCATCCACACGCC	786
Db	842	TCCCTCATGAAGTCCCTGGCCCCATGGGGGGCGCACCGTCTATCTGCACCATCCACACGCC	901
Qy	787	AGTGCCAAGCTCTTTGAGATGTTTGAACAAGCTCTACATCTCTGAGCCAGGGTCAAGTGCATC	846
Db	902	AGTGCCAAGCTCTTTGAGATGTTTGAACAAGCTCTACATCTCTGAGCCAGGGCAATGCATC	961
Qy	847	TTCAAGGAGTGGTACCACCACTGATCCCTATCTTAAAGGACATCGGCTTGCATTTGCCCC	906
Db	962	TTCAAGGAGTGGTATTACCAACCTGATTCCTCTATCTGAAGGGGTGTGGCTTGCATGTCCC	1021
Qy	907	ACCTACCACAACCCGGCTGACTTCACTCATGAGTGGCCCTCTGGCGAGTATGGAGACTG	966
Db	1022	ACCTACCACAACCCGGCTGACTTCACTTATGAGTGGCCCTCTGGAGAGTATGGAGACTG	1081
Qy	967	AAACCCCATGTTGTACGGGCTGTGCAGAAATGGGCTGTGGCTATTTGGCTGAGAAAGAGC	1026
Db	1082	AAACCCCATGTTGTACGGGCTGTGCAGAAATGGGCTTTTGACACCATGGCCGAGAAAGAGC	1141
Qy	1027	AGCCCTGAGAAAGACAGGTTCCCTGCCCCCATGCCCTCTTTGCTCTCGGAAGTGGATCCC	1086
Db	1142	AGCCCTGAGAAAGTACGTTCCCTGCCCCCATGCCCTCTTTGCTCTCGGAAGTGGATCCC	1201

QY	1087	ATTGAAAGCCACACCTTTGGCCACAGCACCCCTCACACAGTTCTGCATCTCTTCAAGAGG	114
DB	1202	ATTGAAAGCCACACATTGGCCACAGCACCCCTAAACCCAGTTCTGCATCTCTTCAGGAG	1261
QY	1147	ACCTTCCTGTCATCTCCACAGGACACGGTCTTGACCCACCTACGGTTTCATGTGCCACGTG	1206
DB	1262	ACTTTCTTGTCATCTCCACAGGACACGGTCTTGACCCACTTGGGGTTTCATGTTCACATGTG	1321
QY	1207	GTTATTGGCGTGTCTATCTGGGCTCTCTACTGTCATATTGGCGACGATGCCAGCAAGTCT	1266
DB	1322	TTGATGGCGTCTCATTTGGCTCTCTACTGCGACATTTGGCGATGATGCCAGCAAGTC	1381
QY	1267	TTCAACAACACGGCTGCTCTTCTTCTCCATGCTGTTCCTCATTTGTCGCCGCCCTCATG	1326
DB	1382	TTCAACAACACCGGCTTCTCTTCTTTCATGCTCTTCTCATGTTTGAGCGCTCTCATG	1441
QY	1327	CCAACTGTGCTACCTTCCCTTTAGAGATGGCGGTCTTCATGAGGGAGACCTCAACTAC	1386
DB	1442	CCGACCGTGTACCTTTTCCCTTTAGAGATGGCGGTCTTCATGAGGGAGACACTTTGAAGTAC	1501
QY	1387	TGTTACAGCCTCAAGCGTATTACCTGGCCAAAGACCATGGCTGACGTGCCCTTTCAGGTG	1446
DB	1502	TGGTATTACTTCAAGCGTATTACTTGGCCAAAGACCATGGCTGATGTGCCCTTCCAGGTG	1561
QY	1447	GTGTGTCGGTGTCTACTGCAGCATTTGTACTGGATGACGGGCCACGCCGTGTGAGACC	1506
DB	1562	GTATGCCACGTGGTGTACTGCAGCATCTGTACTGGATGACAGGCCACGACCGAGACC	1621
QY	1507	AGCGCTTCTGCTCTTCTCAGCCCTGGCCACCGCCACCGCCTTTGGTGGCCCAATCTTTTG	1566
DB	1622	AGCGCTTCTGCTCTTCTCAGCCCTTGGCCATAGCCACCGCCTTAGTTGGCCCAATCTTTTG	1681
QY	1567	GGGCTGTGATCGGAGCTGCTTCCAACTCCCTACAGGTGGCCACTTTTGTGGGCCCAAGTT	1626
DB	1682	GGACTCTTGATGGCGTGCCTCTACCTCTCTTGAGGTGGCCACTTTTGTGGGCCCAAGTT	1741
QY	1627	ACGCCCATCCCTGTCTCTTGTCTCCGGCTTCTTTTCTACAGTTCAGAGCATCAAGCCACT	1686
DB	1742	ACTGCCATCCCTGTCTCTTGTCTCCGGCTTCTTCTGTCAGTTTCAAGACCATCCCCACT	1801
QY	1687	TACCTGCAATGGAGCTCCTATCTCTCCTATGTCAGGTATGGCTTTGAGGTGTGATCTCTG	1746
DB	1802	TACTTTGCAATGGAGCTCCTACCTCTCCTATGTTAGGTATGGCTTTGAGGTCTGATCTCTG	1861
QY	1747	ACGATCTATGGCATGGAGCGAGGACCTGCACATGTTTAGAGGAACGCTGCCCGTTCCTGG	1806
DB	1862	ACCATCTATGGTATGGAGCGAGGACACTGACCTGCTTTAGATGAACAGTGCCCATCTCCGG	1921
QY	1807	GAGCCACAGACATCTCCGAGCGCTGGATGTGGAGGATGCCAGCTCTACATGGACTTC	1866
DB	1922	GACCCAAATCATTTTTCGTGAGCTGATGATAGAGGAGGCCAACTCTACATGGACTTC	1981
QY	1867	CTGGTCTTGGGCATCTTCTTCTTAGCCCTCGGGTGTGGCGTACCTTTGCTGCGTGTAC	1926
DB	1982	CTGGTCTTGGGCATTTTCTTCTTGGCCCTCGGCTGCTGGCATACCTTTGCTCTCCGTAC	2041
QY	1927	CGGTCGAATCAGAGATAGAGCTTTGCCCCAGCCCTGTACCCCGCTGCGAGCAGGA	1986
DB	2042	CGGTCGAATCAGAGATAGAGCTTGGCTTCTACCTCTTC-----TGGGAGGA	2091
QY	1987	AGCCCCAGTCCCAGCCCTTTTGGGACTGTTTTAACCTTATAGACTTTGGGCACCTGGTCTCT	2046
DB	2092	AGGCCCCAGTCCCAGCCCTTTTGGGCTGTTTTAACTTGT- - ACTTGGGCATTTGTTGCT	2149
QY	2047	GGCGGGCTATCTCTCTCCTCTGGCTCTCCACAGGCTGGCTGTG- - -GGACTGGCT	2103
DB	2150	GGTGCAGCTGCCCTCTCTCTCCCTCGGCTGTACTCAGGCTGGCTGTGGGCTGTGTGCT	2209
QY	2104	CCAGCGCTGGGCTCTGGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCTTCC	2163
DB	2210	TCCAGTCTCAGCCCCGAGGTTAGGGCTCCAGGCCCTCCGCTC- - -AAGGCCAGGAGTCTTCC	2268
QY	2164	CAAGTTGATCGGTTTGTAGCTTTCCTCTCTCTCTCCAAACACCTCGCATGCAAGAACTA	2223

CDS 123..2063
/gene="ABC4"
/codon_start=1
/product="putative white family ABC-transporter"
/protein_id="CAD19779.2"
/db_xref="GI:18496085"
/translation="MAEKALAEVYCGGLGFCGAVAMAVTLEDGAPPPVLTTHLKVENHI
TEARFSLHPKRSADVIEFVELSYVREGPCWRKRGYKTLKCLSGKFCRRELIGLWG
PSGAGKSTFNNILAGYREGKMGGOILVNGRPRRLRFRKMSVIMODDMLPLHLTVNN
AMVSNLKLSEKQEVKELVTEILALGMSCHPTALLSGGRKRLAIALELVNN
PPVWFDEPTSLGDSASCFQVYSLMSLAHGRVICTIHPQSAKLFEMFDKYLISQ
GCIFKGVNLIPIKGLGLHCPYTHNPADFIIEVASGEYGLNPLFRVAVONGILT
MAEKSPKSNELPAPKCPPELDPESHTFATSLITQFCLFRFTLSILRDVLT
HLFMRHLVIGLILYHLIGDDASKVFNNTGFLFSLMLFMAALMPTLFFPLEM
AVFMRHLNWTYKAYLAKTMADVPVQVPCVYVNMVQOPATSRFLPFA
LATALVAQSLGLLGAASTLSQVATFVGPVTAIPVLFSFVFSFKTIPYVLOWS
YLSVRYKPEGLIITIGHMERHLCLEQCPFRDPOIILRELDEVEAKLYMDFLVLG
IFFLAURLALYLIRVYRVSER"
BASE COUNT 675 a 1078 c 917 g 850 t
ORIGIN

Query Match 61.5%; Score 2125; DB 10; Length 3520;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 2776; Conservative 0; Mismatches 565; Indels 112; Gaps 20;

QY 7 ATGGCGAGAGCGGTGGAGCCGTGGGCTGTGGACTAGGGCCGGGGCTGTGGCCATG 66
DB 123 ATGGCGAGAGGCATTGGAGCGGTGGGCTGTGGACTGGGCCCTGGGGCTGTGGCCATG 182

QY 67 GCGGTGACCTGAGGACGGGGGGAACCCCTGTGCTGACACGACACCTGAGAGAGGTG 126
DB 183 GCTGTGACCTGAGGACGGGGGAGAGCCCTGTGCTGACACGACACCTTAAGAGGTG 242

QY 127 GAGAACCATATCAGAGACAGAGCGCTTCTCCACCTGCCCCAGCGCTCAGCGCTGGAC 186
DB 243 GAGAACCATATCAGAGACAGAGCGCTTCTCCACCTGCCCCAGCGCTCAGCGCTGGAC 302

QY 187 ATCGAGTTGCGAGCTGCTTATTCGGTGGGGAGGGGGCCCTGTGGCGCAAAAGGGGT 246
DB 303 ATCGAGTTGCGAGCTGCTTATTCGGTGGGGAGGGGGCCCTGTGGCGCAAAAGGGGT 362

QY 247 TATAAGACCTTCTCAAGTGCCTCTCAGGTAAATCTGCGCGGGAGCTGATGGCATC 306
DB 363 TATAAGACCTTCTCAAGTGCCTCTCAGGTAAATCTGCGCGGGAGCTGATGGCATC 422

QY 307 ATGGGCCCCCTCAGGGCTGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGGAG 366
DB 423 ATGGGCCCCCTCAGGGCTGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGGAG 482

QY 367 TCTGGAATGAAGGGGAGATCTCTGTTAATGGAAGCCACGCGGAGCTGAGGACCTTCCCG 426
DB 483 TCGGGGATGAAGGGGAGATCTCTGTTAATGGAAGCCACGCGGAGCTGAGGACCTTCCCG 542

QY 427 AAGATGCTCTGCTACATCATGCAAGATGACATGCTGCTGCCACCTCAGGTGTTGGAA 486
DB 543 AAGATGCTCTGCTACATCATGCAAGATGACATGCTGCTGCCACCTCAGGTGTTGGAA 602

QY 487 GCCATGATGGTCTCTGCTTAACCTGAATCTTACTGAGAATCCCGATGTGAAAACGATCTC 546
DB 603 GCCATGATGGTCTCTGCTTAACCTGAATCTTACTGAGAATCCCGATGTGAAAACGATCTC 662

QY 547 GTGACAGATCTTGACGGCATGGGCCCTGATGCTGCTGCCACAGGAGACAGCCCTG 606
DB 663 GTGACAGATCTTGACGGCATGGGCCCTGATGCTGCTGCCACAGGAGACAGCCCTG 722

QY 607 CTCTCTGGGGGAGAGAGCGCTGTGGCCATCGCCCTGGAGCTGTGCAACACCGCCCT 666
DB 723 CTCTCTGGGGGAGAGAGCGCTGTGGCCATCGCCCTGGAGCTGTGCAACACCGCCCT 782

QY 667 GTCATGTTCTTTGATGAGCCACCACTGCTGCTGATAGCGCCCTTGTGTTTCCAAAGTGGT 726
DB 783 GTCATGTTCTTTGATGAGCCACCACTGCTGCTGATAGCGCCCTTGTGTTTCCAAAGTGGT 842

QY 727 TCCTCATGAAGTCCCTGGCACAGGGGGCCGTACCATCATCTGACCATCCACGCCC 786
DB 843 TCCTCATGAAGTCCCTGGCCATGGGGCCGACCGTCATCTGCACCATCCACGCCC 902

QY 787 AGTGCCAAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTTGAGCCAGGGTCAAGTCA 846
DB 903 AGTGCCAAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTTGAGCCAGGGCAATGCATC 962

QY 847 TTCAAAGGAGTGTGTACCAACCTGATCCCTATCTAAAGGAGCTCCGGCTTGCATTCGCC 906
DB 963 TTCAAAGGAGTGTGTACCAACCTGATCCCTATCTAAAGGAGCTCCGGCTTGCATTCGCC 1022

QY 907 ACCTACCAACACCCGGCTGACTTTCATCATCGAGGTGGCCCTCTGGCCAGTATGAGACCTG 966
DB 1023 ACCTACCAACACCCGGCTGACTTTCATCATGAGGTGGCCCTCTGGAGAGTATGAGACCTG 1082

QY 967 AACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTGTGGCTATGCTGAGAGAGAGC 1026
DB 1083 AACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTTTCGACCATGGCCGAGAGAGAGC 1142

QY 1027 AGCCCTGAGAAGAACGAGGTCCCTGCCCATGCCCTCTGTCTCTCCGGAAGTGATCCC 1086
DB 1143 AGCCCTGAGAAGAACGAGGTCCCTGCCCATGCCCTCTGTCTCTCCGGAAGTGATCCC 1202

QY 1087 ATTGAAGCCACACCTTTGCCACGACCCCTCACACAGTTCCTGCACTCTCTTCAAGAGG 1146
DB 1203 ATTGAAGCCACACATTTGCCACGACCCCTAACCCAGTTCCTGCACTCTCTTCAAGAGG 1262

QY 1147 ACCTTCTCTCCATCTCAGGGACAGGGTCCCTGACCCACCTAGCGTTCATGTCCTCCAGTG 1206
DB 1263 ACCTTCTCTCCATCTCAGGGACAGGGTCCCTGACCCACCTAGCGTTCATGTCCTCCAGTG 1322

QY 1207 GTTATTGGGTGCTCATCGCCCTCTTACTGTCATATTGGGACGATGCCACCAAGGTC 1266
DB 1323 TTGATCGGGGTCTCTCAATGGCCCTCTTACTGTCACATTTGGCGGTTCTATGTCACATGTG 1382

QY 1267 TTCAACAACACCGGGTGCCTCTTCTTCTCATGCTCTTCTCATGTTGCGCGCCCTCATG 1326
DB 1383 TTCAACAACACCGGGTGCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1442

QY 1327 CCAACTGTCTCACCTTCCCTTTAGAGATGGCGGTCTTCTATGAGGAGGACCTCAACTAC 1386
DB 1443 CCGACCGTGTCTCACCTTCCCTTTAGAGATGGCGGTCTTCTATGAGGAGGACCTTGAACATC 1502

QY 1387 TGTACAGCTCTCAAGCGTATTACCTGGCCCAAGACCATGGCTGCGCTGCGCTTTCAGGTG 1446
DB 1503 TGTATCTCTCAAGCGCTATTACTTGGCCCAAGACCATGGCTGATGCTGCTTCCAGGTG 1562

QY 1447 GTGTGTCGGTGTCTTACTGACGATTTGTACTGGATGACGGGCGACCGCGCTGAGACC 1506
DB 1563 GTATGCCAGTGTGTACTGACGATCTGTACTGGATGACAGGCGAGCGAGAGC 1622

QY 1507 AGCGGCTTCTCTCTCTTCTCAGCCCTGGCCACCGCCCTTGTGTGGCCCAATCTTTG 1566
DB 1623 AGCGGCTTCTCTCTCTTCTCAGCCCTGGCCATAGCCACCGCCCTTGTGTGGCCCAATCTTTG 1682

QY 1567 GGCTCTGCTGAGGAGTCTTCTTCCAACTCCCTACAGGTGGCCACCTTTTGTGGGCCAGTT 1626
DB 1683 GGACTCTTGTATGGGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1742

QY 1627 ACGGCATCCCTGCTCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1686
DB 1743 ACTGCCATCCCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1802

QY 1687 TACCTGCAATGAGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1746
DB 1803 TACTTCAATGAGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1862

QY 1747 ACCATCTATGGCATGAGGAGGAGACCTGACATGTTTATAGAGAACGCTGCCCGTTCCTCCG 1806
DB 1863 ACCATCTATGGCATGAGGAGGAGACCTGACCTGCTTAGATGAACAGTGCCTTCCCG 1922

QY 1807 GAGCCACAGAGCATCTCCGAGCGGTGGATGTGGAGATGCCAAGCTCTACATGGACTTC 1866

Center clone name: CMB9-93G21
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 130262 bases at least Q40
Consensus quality: 136359 bases at least Q30
Consensus quality: 143933 bases at least Q20
Insert size: 151905; sum-of-ctgigs
Quality coverage: 8.32x in Q20 bases; sum-of-ctgigs

NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1
7883 contig of 7883 bp in length
14891 contig of 6908 bp in length
22645 contig of 7654 bp in length
29770 contig of 6925 bp in length
36703 contig of 6933 bp in length
43057 contig of 6254 bp in length
48313 contig of 5156 bp in length
53404 contig of 4991 bp in length
57461 contig of 3957 bp in length
61460 contig of 3899 bp in length
64861 contig of 3301 bp in length
68047 contig of 3086 bp in length
71295 contig of 3148 bp in length
74584 contig of 3188 bp in length
77136 contig of 2453 bp in length
81080 contig of 3844 bp in length
84386 contig of 3206 bp in length
88487 contig of 3206 bp in length
90954 contig of 3206 bp in length
93473 contig of 3206 bp in length
97782 contig of 2035 bp in length
100347 contig of 2465 bp in length
102232 contig of 1785 bp in length
102333 contig of 2700 bp in length
105133 contig of 3301 bp in length
108534 contig of 1776 bp in length
110410 contig of 2275 bp in length
112684 contig of 1592 bp in length
114477 contig of 2071 bp in length
116648 contig of 1822 bp in length
118570 contig of 2154 bp in length
120824 contig of 1746 bp in length
122569 contig of 2229 bp in length
124898 contig of 2605 bp in length
126570 contig of 1648 bp in length
127603 contig of 1324 bp in length
129351 contig of 1662 bp in length
130775 contig of 1353 bp in length
130876 contig of 1327 bp in length
132638 contig of 1280 bp in length
133990 contig of 1361 bp in length
134091 contig of 1432 bp in length
135417 contig of 1216 bp in length
136797 contig of 1155 bp in length
138258 contig of 1308 bp in length
138359 contig of 1796 bp in length
139790 contig of 1407 bp in length
141106 contig of 1812 bp in length
142361 contig of 1405 bp in length
143769 contig of 1076 bp in length
145665 contig of 1470 bp in length
147172 contig of 1224 bp in length
149084 contig of 1048 bp in length
149185 contig of 1048 bp in length
150589 contig of 1048 bp in length
151765 contig of 1048 bp in length
151866 contig of 1048 bp in length
153436 contig of 1048 bp in length
154659 contig of 1048 bp in length
155807 contig of 1048 bp in length

155908 157405 contig of 1498 bp in length.
NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1
7883 contig of 7883 bp in length
7884 7983: gap of 100 bp
14891 14891: gap of 100 bp
14892 14991: gap of 100 bp
22645 22645: gap of 100 bp
22646 22745: gap of 100 bp
22746 29670: gap of 100 bp
29671 29770: gap of 100 bp
29771 36703: gap of 100 bp
36704 36803: gap of 100 bp
36804 43057: gap of 100 bp
43058 43157: gap of 100 bp
43158 48313: gap of 100 bp
48314 48413: gap of 100 bp
48414 53404: gap of 100 bp
53405 53504: gap of 100 bp
53505 57461: gap of 100 bp
57462 57561: gap of 100 bp
57562 61460: gap of 100 bp
61461 61560: gap of 100 bp
61561 64861: gap of 100 bp
64862 64961: gap of 100 bp
64962 68047: gap of 100 bp
68048 68147: gap of 100 bp
68148 71295: gap of 100 bp
71296 71395: gap of 100 bp
71396 74583: gap of 100 bp
74584 74683: gap of 100 bp
74684 77136: gap of 100 bp
77137 77236: gap of 100 bp
77237 81080: gap of 100 bp
81081 81180: gap of 100 bp
81181 84386: gap of 100 bp
84387 84486: gap of 100 bp
84487 87351: gap of 100 bp
87352 87451: gap of 100 bp
87452 88441: gap of 100 bp
88442 88941: gap of 100 bp
88942 90853: gap of 100 bp
90854 90953: gap of 100 bp
90954 93372: gap of 100 bp
93373 93472: gap of 100 bp
93473 95647: gap of 100 bp
95648 95747: gap of 100 bp
95748 97782: gap of 100 bp
97783 97882: gap of 100 bp
97883 100347: gap of 100 bp
100348 100447: gap of 100 bp
100448 102232: gap of 100 bp
102233 102332: gap of 100 bp
102333 105032: gap of 100 bp
105033 105132: gap of 100 bp
105133 108433: gap of 100 bp
108434 108533: gap of 100 bp
108534 110309: gap of 100 bp
110310 110409: gap of 100 bp
110410 112684: gap of 100 bp
112685 112784: gap of 100 bp
112785 114376: gap of 100 bp
114377 114476: gap of 100 bp
114477 116547: gap of 100 bp
116548 116647: gap of 100 bp
116648 118469: gap of 100 bp
118470 118569: gap of 100 bp

Db	13097	CCACCACACAAGTGTGTTGTAAACTGGGCTGCTATAAAGGTGGAGTTCACAGGCGCTGGGCC	13151
Qy	2401	TGCTGGAGTCCACTGGAAGTCCCACTTATGTGATGTTGAAATGACAGGAAGGACTCTGGA	2460
Db	13157	TGCTGGAGTCCACTGGAAGTCCCACTTATGCGATGTTGAAATGACAGGAAGGACTCTGGA	13216
Qy	2461	AGTCTCTTCCCTCCCTCTCTCTCTCTCTCCACCCCTAGACCCCTGGCTGACTTGGACAATCT	2520
Db	13217	AGTCTCTTCCCTCTCTCTCTCTCTCTCCACCCCTAGACCCCTGGCTGACTTGGACAATCT	13276
Qy	2521	GCCAGACAGAGCTGGGTTTCTGTCTAGGTACCACTCCCAATCTCTGGGATTTGGAGA	2580
Db	13277	GCCAGACAGAGCTGGGTTTCTGTCTAGGTACCACTCCCAATCTCTGGGATTTGGAGA	13336
Qy	2581	GGCTCTGGGCTGTGGGATGCCCATCCCCCTCCCATCACTTTGGTGGGGCAGGGCT	2640
Db	13337	GGCTCTGGGCTGTGGGATGCCCATCCCCCTCCCATCACTTTGGTGGGGCAGGGCT	13396
Qy	2641	GGTGGCACCTGTGCAATAATGTCTGTGTTTCTCTCCCACTGCCACTGGAATCGAGAAT	2700
Db	13397	GGTGGCACCTGTGCAATAATGTCTGTGTTTCTCTCCCACTGCCACTGGAATCGAGAAT	13456
Qy	2701	GCATTTATTTCTGGGGGGGGTGTAGTGGGGGAAGACCAACCCCTCTTTCTCGCTGCC	2760
Db	13457	GCATTTATTTCTGGGGGGGGTGTAGTGGGGGAAGACCAACCCCTCTTTCTCGCTGCC	13516
Qy	2761	CTAACCGATGCACGGTCTCGTGATGCTCCCTCCCTCTCGGAGTGACAGGCACATACATG	2820
Db	13517	CTAACCGATGCACGGTCTCGTGATGCTCCCTCCCTCTCGGAGTGACAGGCACATACATG	13576
Qy	2821	AGAACAGGCCATCTACGCCCTACACACTTGGCCATCCCTTACAGCACAGAGGAAGTGTAT	2880
Db	13577	AGAACAGGCCATCTACGCCCTACACACTTGGCCATCCCTTACAGCACAGAGGAAGTGTAT	13636
Qy	2881	GGTGGCATGCTGTGTGGGGGTGCTGTGTGGAGGACAGTGCCCAACCTCTCTCTGGGA	2940
Db	13637	GGTGGCATGCTGTGTGGGGGTGCTGTGTGGAGGACAGTGCCCAACCTCTCTCTGGGA	13696
Qy	2941	TCCCATGTTGGACACTTAAGGATAAGCTGTGTGTGCCACGGGTGCTTACAGGAACATGC	3000
Db	13697	TCCCATGTTGGACACTTAAGGATAAGCTGTGTGTGCCACGGGTGCTTACAGGAACATGC	13756
Qy	3001	AGGTGTCTACCCCAAGTCTTCCCTCTCCCAAGCCAGGGTGGCACAGGGCAGTACATC	3060
Db	13757	AGGTGTCTACCCCAAGTCTTCCCTCTCCCAAGCCAGGGTGGCACAGGGCAGTACATC	13816
Qy	3061	CTGTGGATTCAGGAACACACAGGACAAACACAGGGGATAGTGTGGCTTGGCCACTGCG	3120
Db	13817	CTGTGGATTCAGGAACACACAGGACAAACACAGGGGATAGTGTGGCTTGGCCACTGCG	13876
Qy	3121	CACCCAGGGCCCTTTTGTGTCTCCATGCTGGCATCTTCACTCCCTTACCCCTTCCCCA	3180
Db	13877	CACCCAGGGCCCTTTTGTGTCTCCATGCTGGCATCTTCACTCCCTTACCCCTTCCCCA	13936
Qy	3181	GCCACTGCTGCTCATTTCAAACTTCTGTCCATGTCCCTCCCACTGTTCTATCAGCAGTGG	3240
Db	13937	GCCACTGCTGCTCATTTCAAACTTCTGTCCATGTCCCTCCCACTGTTCTATCAGCAGTGG	13996
Qy	3241	CCCTCTGGGCATCAGAACAGGCTGCCCTGGGCACCAAGTGGGACACACTCAGAGCATGT	3300
Db	13997	CCCTCTGGGCATCAGAACAGGCTGCCCTGGGCACCAAGTGGGACACACTCAGAGCATGT	14056
Qy	3301	CTGGCTTCTCTGGTGGGTCCAGGCTCATTTCTGTTCTGATTTCCCTTCCCTCCAGGGCTCA	3360
Db	14057	CTGGCTTCTCTGGTGGGTCCAGGCTCATTTCTGTTCTGATTTCCCTTCCCTCCAGGGCTCA	14116
Qy	3361	TTTTTCCCTCTTTTCTGTACACATCCCTGTCTTACCTCTCTCACCTGCCACAGATTTCT	3420
Db	14117	TTTTTCCCTCTTTTCTGTACACATCCCTGTCTTACCTCTCTCACCTGCCACAGATTTCT	14176
Qy	3421	TCCTATCACACAGGGATGCCAGTTGATTTGTGGG	3455
Db	14177	TCCTATCACACAGGGATGCCAGTTGATTTGTGGG	14211


```

Db      1567 GCAGACACACTCAGACGATGTCGGCTTCTCCTGGTGGGTCCAGGCTCATCTGCTCTCTGA 1626
QY      3340 TTTCCCTCCCCCAGGGCTCATTTTCCCTTTTCTCTGTACACATCCCTGCTACCTCC 3399
Db      1627 TTTCCCTCCCCCAGGGCTCATTTTCCCTTTTCTCTGTACACATCCCTGCTACCTCC 1686
QY      3400 TCTCACCTTCCACAGATTTCTCTATACACAGGATGCCAGTTGATTTGGG 3455
Db      1687 TCTCACCTTCCACAGATTTCTCTATACACAGGATGCCAGTTGATTTGGG 1742

RESULT 9
AP002956
LOCUS   Homo sapiens genomic DNA, chromosome 11q clone:R1105h09, complete
DEFINITION
ACCESSION AP002956
VERSION   1
KEYWORDS HTG.
SOURCE    Homo sapiens DNA, clone:R1105h09.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS  Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE    Homo sapiens 219,574 genomic DNA of 11q
JOURNAL  Published Only in Database (2002)
REFERENCE
2 (bases 1 to 219574)
AUTHORS  Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE    Direct Submission
JOURNAL  Submitted (24-NOV-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gs.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Overlapping Clones: pD1A4, PQ38G8
Cytogenetic Position: 21q22.1, region:D21S226-AML
STS Markers (ePCR): stSG53747, SHGC-16045, D11S2450, SHGC-7143,
WI-16056.
stSG31054, sts-X57110, WI-8978, GDB:198117, stSG29284, SHGC-130657,
D11S1941E, WI-14589, stSG60191, stSG50535, A002048, stSG26946,
stSG41092,
TIGR-A005D38, Baa7902, D11S869E, RH80030
Contamination: none detected
non-AGCT bases: none
Additional author information
Arai,Y., Kubo,T.,Ohki,M.
National Cancer Center Research Institute
Cancer Genomic Division
5-1-1, Tsukiji, Chuo-ku, Tokyo, JAPAN
zip: 104-0045
phone: +81-3-3542-2511 ex4752, fax: +81-3-3542-0688 e-mail:
yaraie@cc.riken.go.jp.
FEATURES
source
1. .219574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="R1105h09"
BASE COUNT 54883 a 50356 c 53844 g 60491 t
ORIGIN

Query Match 50.1%; Score 1729.4; DB 9; Length 219574;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1753; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

QY      1681 CCACCTTACCTGCAATGGAGCTCCTATCTCTCTATGTCAGGATGCTTGGAGGTGTG 1740
Db      30111 CCCTGTGCTGTATGCCATCTGATATCCCTGCCTAGGATGGCTTTGAGGGTGTG 30170

```

```

QY      1741 ATCTGACGATCTATGGCATGGAGGAGACCTTGACATGTTTAGAGGAACGCTGCCCG 1800
Db      30171 ATCTGACGATCTATGGCATGGAGGAGACCTTGACATGTTTAGAGGAACGCTGCCCG 30230
QY      1801 TTCCGGGAGCACAGAGCATCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATG 1860
Db      30231 TTCCGGGAGCACAGAGCATCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATG 30290
QY      1861 GACTTCTGCTGTGGGCATCTTCTTCCCTAGCCCTGCGGCTGCTGGCTACCTTGTGCTG 1920
Db      30291 GACTTCTGCTGTGGGCATCTTCTTCCCTAGCCCTGCGGCTGCTGGCTACCTTGTGCTG 30350
QY      1921 CGTTACCGGCTCAAGTCAGAGATAGAGCTTGCCTCCAGCCCTGTATCCCGAGCCCTGCA 1980
Db      30351 CGTTACCGGCTCAAGTCAGAGATAGAGCTTGCCTCCAGCCCTGTATCCCGAGCCCTGCA 30409
QY      1981 GCAGGAAGCCCCCAGTCCAGACCTTTTGGACCTGTTTAACTTTATAGACTTTGGGCAC 2040
Db      30410 GCAGGAAGCCCCCAGTCCAGACCTTTTGGACCTGTTTAACTTTATAGACTTTGGGCAC 30469
QY      2041 GTTCTTGGGGGGCTATCTCTCTCTTGGCTCTCCACAGGCTGGCTGTCCGACTGC 2100
Db      30470 GTTCTTGGGGGGCTATCTCTCTCTTGGCTCTCCACAGGCTGGCTGTCCGACTGC 30529
QY      2101 GCTCCAGGCTGGGCTCTGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT 2160
Db      30530 GCTCCAGGCTGGGCTCTGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT 30589
QY      2161 TCCCAAGTTGATCGGTTTGTAGCTTCCCTTACTCTCTCCACAGCTGCATGCCAAGA 2220
Db      30590 TCCCAAGTTGATCGGTTTGTAGCTTCCCTTACTCTCTCCACAGCTGCATGCCAAGA 30649
QY      2221 CTACTGGGAGGCTGCTCCCTCTCTGCTCCATGGACACCTCTCTGCTCTGCTGGG 2280
Db      30650 CTACTGGGAGGCTGCTCCCTCTCTGCTCCATGGACACCTCTCTGCTCTGCTGGG 30709
QY      2281 AGCCCTAGGCTCTCTAGGGCCCCACTTACAACCTGACCAAGTGGCCCCCTCTGGGGTCC 2340
Db      30710 AGCCCTAGGCTCTCTAGGGCCCCACTTACAACCTGACCAAGTGGCCCCCTCTGGGGTCC 30769
QY      2341 CCACCACACAGTGTGTTGTAACCTGGCTGCTATAGGTTGGAGTTCCAGGGCTGGGCC 2400
Db      30770 CCACCACACAGTGTGTTGTAACCTGGCTGCTATAGGTTGGAGTTCCAGGGCTGGGCC 30829
QY      2401 TGTGTGAGTCCACTGGAAGTCCCATTTATGATGTTGAAATGGACAGGAAGGACTCTGGA 2460
Db      30830 TGTGTGAGTCCACTGGAAGTCCCATTTATGATGTTGAAATGGACAGGAAGGACTCTGGA 30889
QY      2461 AGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520
Db      30890 AGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 30949
QY      2521 GCCAGGACAGAGCTGGGTTTCTGCTAGTTCACCACTCCCAATCCTGGGATTTGAGA 2580
Db      30950 GCCAGGACAGAGCTGGGTTTCTGCTAGTTCACCACTCCCAATCCTGGGATTTGAGA 31009
QY      2581 GGCCTGGGCTGTGGATGCCCATCCCTCTCCCATCATCTTTTGGTGGGGGAGGGCCCT 2640
Db      31010 GGCCTGGGCTGTGGATGCCCATCCCTCTCCCATCATCTTTTGGTGGGGGAGGGCCCT 31069
QY      2641 GGTGGCACCTGTGCAATAATGCTGTGTGTTTCTCTCCCACTGCCACTTGGAACTGGAGAAT 2700
Db      31070 GGTGGCACCTGTGCAATAATGCTGTGTGTTTCTCTCCCACTGCCACTTGGAACTGGAGAAT 31129
QY      2701 GCATTATTATCTGGGGGGGGTGTGGGGGAGACCCCAACCTCTCTCTCTCTCTCTCTCTCT 2760
Db      31130 GCATTATTATCTGGGGGGGGTGTGGGGGAGACCCCAACCTCTCTCTCTCTCTCTCTCTCT 31189
QY      2761 CTAACGCATCCAGGCTCTCGTGTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2820
Db      31190 CTAACGCATCCAGGCTCTCGTGTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 31249
QY      2821 AGAACAGGCCATCTCAGCCCTACACACTTGGCCATCCCTTACACACAGAGGAAGAGTGTAT 2880

```


85736	89553	contig of	2818	bp	in	length
88654	92373	contig of	3720	bp	in	length
92474	95785	contig of	3312	bp	in	length
95886	99796	contig of	3911	bp	in	length
99897	103432	contig of	3536	bp	in	length
103533	107518	contig of	3986	bp	in	length
107619	110312	contig of	2694	bp	in	length
110413	112020	contig of	1608	bp	in	length
112121	114398	contig of	2278	bp	in	length
114499	117281	contig of	2783	bp	in	length
117382	120102	contig of	2721	bp	in	length
120203	122509	contig of	2307	bp	in	length
122610	125301	contig of	2692	bp	in	length
125402	127610	contig of	2209	bp	in	length
127711	130017	contig of	2307	bp	in	length
130118	132370	contig of	2253	bp	in	length
132471	134840	contig of	2370	bp	in	length
134941	135021	contig of	81	bp	in	length
135122	136660	contig of	1539	bp	in	length
136761	138314	contig of	1554	bp	in	length
138415	139654	contig of	1240	bp	in	length
139755	140914	contig of	1160	bp	in	length
141015	142197	contig of	1183	bp	in	length
142298	143507	contig of	1210	bp	in	length
143608	144794	contig of	1187	bp	in	length

*	110313	110412:	gap of	100 bp
*	110413	112020:	contig of	1608 bp
*	110421	112120:	gap of	100 bp
*	111201	112120:	gap of	100 bp
*	111211	114398:	contig of	2278 bp
*	114399	114698:	gap of	100 bp
*	114499	117281:	contig of	2783 bp
*	117282	117381:	gap of	100 bp
*	117382	120102:	contig of	2721 bp
*	120103	120202:	gap of	100 bp
*	120203	122509:	contig of	2307 bp
*	122510	122609:	gap of	100 bp
*	122610	125301:	contig of	2692 bp
*	125302	125401:	gap of	100 bp
*	125402	127610:	contig of	2209 bp
*	127611	127710:	gap of	100 bp
*	127711	130017:	contig of	2307 bp
*	130018	130117:	gap of	100 bp
*	130118	132370:	contig of	2253 bp
*	132371	132470:	gap of	100 bp
*	132471	134840:	contig of	2370 bp
*	134841	134940:	gap of	100 bp
*	134941	135021:	contig of	81 bp
*	135022	135121:	gap of	100 bp
*	135122	136660:	contig of	1539 bp
*	136661	136760:	gap of	100 bp
*	136761	138334:	contig of	1554 bp
*	138315	138414:	gap of	100 bp
*	138415	139654:	contig of	1240 bp
*	139655	139754:	gap of	100 bp
*	139755	140914:	contig of	1160 bp
*	140915	141014:	gap of	100 bp
*	141015	142197:	contig of	1883 bp
*	142198	142297:	gap of	100 bp
*	142298	143507:	contig of	1210 bp
*	143508	143607:	gap of	100 bp
*	143608	144794:	contig of	1187 bp

	Query Match	49.7%	Score 1717.4	DB 2	Length 144794
	Best Local Similarity	98.7%	Pred. No. 0		
	Matches 1752	Conservative	0	Mismatches	21
				Indels	2
				Gaps	2
Qy	1681	CCCACTTACCTGC	CAATGAGCCTCCTATCTCTCCCTATGTCAGGATATGCTTGTAGAGGGTGTG	1740	
Db	10637	CCCCCTTCTGTGTATCCCATCTGATATCCCTGCCTAGCTAGGATGCTTTGAGGGTGTG		10691	
Qy	1741	ATCCTGACGATCTATGCGATGGAGCGAGGACACCTGACATGTTTAGAGGAACGCTGCCCG		1800	

Db	10697	ATCTGACGATCTATGGCATATGGAGCGAGAGACCTTGACATGTTTATAGGAACGCTGCCCC	10756
Qy	1801	TTCCGGAGGCCACAGAGCATCTCCGAGCGCTTGGATGTGGAGATGCCAAGCTCTACATG	1860
Db	10757	TTCCGGAGGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGATGCCAAGCTCTACATG	10816
Qy	1861	GACTTCCTGGTCTTTGGGCATCTTCTTCTTAGCCCTTGCCTGTCTGGCTTGTGGCTACCTTTGTGCTG	1920
Db	10817	GACTTCCTGGTCTTTGGGCATCTTCTTCTTAGCCCTTGCCTGTCTGGCTTGTGGCTACCTTTGTGCTG	10876
Qy	1921	CGTTACCGGCTCAAGTTCAGAGATAGAGGCTTTGCCCGAGCCTGTACCCAGCCCTCTGCA	1980
Db	10877	CGTTACCGGCTCAAGTTCAGAGATAGAGGCTTTGCCCGAGCCTGTACCCAGCCCTCTGCA	10936
Qy	1981	GCAGGAAGCCCCCAGTCCCGAGCCCTTTGGGACTGTTTTAACTTATAGACTTTGGGCACTG	2040
Db	10937	GCAGGAAGCCCCCAGTCCCGAGCCCTTTGGGACTGTTTTAACTTATAGACTTTGGGCACTG	10996
Qy	2041	GTTCCTGGGGGCTATCCCTTCTTCTTCCCTTGGCTCCTCCACAGGCTGCTCTCGGACTGC	2100
Db	10997	GTTCCTGGGGGCTATCCCTTCTTCTTCCCTTGGCTCCTCCACAGGCTGCTCTCGGACTGC	11056
Db	2101	GCCTCCAGCGCTGGGCTCTGGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT	2160
Db	11057	GCCTCCAGCGCTGGGCTCTGGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT	11116
Qy	2161	TTCCCAAGTTGATCGGGTTTGTAGCTTCTCCTTACTCTCTCAACACCTGCATGCAAGA	2220
Db	11117	TTCCCAAGTTGATCGGGTTTGTAGCTTCTCCTTACTCTCTCAACACCTGCATGCAAGA	11176
Qy	2221	CTACTGGGAGGCTGTGCCTCCTCCTGCCATGSCACCCCTCCTCTGTCTGCCTGGG	2280
Db	11177	CTACTGGGAGGCTGTGCCTCCTCCTGCCATGSCACCCCTCCTCTGTCTGCCTGGG	11236
Qy	2281	AGCCCTTAGGCTCTTAGGGCCCCACTTACAACTGACCAAAAGTGGCCCCCTCTGGGGTCC	2340
Db	11237	AGCCCTTAGGCTCTTAGGGCCCCACTTACAACTGACCAAAAGTGGCCCCCTCTGGGGTCC	11296
Qy	2341	CCACCACACAAGTGTGTGTAACCTGGGCTGCTATTAAGGTTCGAGTTCACAGGCTGGGCC	2400
Db	11297	CCACCACACAAGTGTGTGTAACCTGGGCTGCTATTAAGGTTCGAGTTCACAGGCTGGGCC	11355
Qy	2401	TGCTGGAGTCCACTGGAAGTCCCAATTATGGATGTTGAAATGACAGGGAAGGACTCTGGA	2460
Db	11356	TGCTGGAGTCCACTGGAAGTCCCAATTATGGATGTTGAAATGACAGGGAAGGACTCTGGA	11415
Qy	2461	AGTCTCTTCCCTCTCTCTCTCTCCACCCCTAGACCTTGGCTGGCTGACTTGGACAATCT	2520
Db	11416	AGTCTCTTCCCTCTCTCTCTCTCCACCCCTAGACCTTGGCTGGCTGACTTGGACAATCT	11475
Db	2521	GCCAGGACAGAAGCTGGGTTTTCTCTCTAGGTCAACCACTCCCAATCTCGGGATGGAGA	2580
Db	11476	GCCAGGACAGAAGCTGGGTTTTCTCTCTAGGTCAACCACTCCCAATCTCGGGATGGAGA	11535
Qy	2581	GGCCTGGGCTGTGGGATGCCCAATCCCCCTCCCACTCACCTTTGGTGGGGCAGGGCCT	2640
Db	11536	GGCCTGGGCTGTGGGATGCCCAATCCCCCTCCCACTCACCTTTGGTGGGGCAGGGCCT	11595
Qy	2641	GGTGGCACTGTGCAATTAATGTCTGTGTTTCTCTCCCACTTCCCACTGGCACTGGAGAAT	2700
Db	11596	GGTGGCACTGTGCAATTAATGTCTGTGTTTCTCTCCCACTTCCCACTGGCACTGGAGAAT	11655
Qy	2701	GCACTTTATTTCTGGCGGGGTGAGTGGGGAGAGCCCAACCTCTCTTCTCGCTGCC	2760
Db	11656	GCACTTTATTTCTGGCGGGGGTGTGAGTGGGGAGAGCCCAACCTCTCTTCTCGCTGCC	11715
Qy	2761	CTAAGCGATGACGGTCTCGTGATGCTCCCTTCCCTCTCGGAGTGACAGGCACATACATG	2820
Db	11716	CTAAGCGATGACGGTCTCGTGATGCTCCCTTCCCTCTCGGAGTGACAGGCACATACATG	11775
Qy	2821	AGAACAGGCCATCTCAGCCCTTACACATTTGCCATCCCTTACAGCACAGAGGAAGTGAT	2880

Db	11776	AGAACAGGCCATCTCAGCCCTACACACTTGGCCATCCCCTACAGCACACAGGAAGATGAT	11835
Qy	2881	GGTGGCATGCTGGTGGCGGGTGTGGTGGGAGACAGTCCCAACCTCTCTCTGGGGA	2940
Db	11836	GGTGGCATGCTGGTGGCGGGTGTGGTGGGAGACAGTCCCAACCTCTCTCTGGGGA	11895
Qy	2941	TCCCATGTTGGAGACTCTAAGGATAAGGCTGGTGTGCCCGGAGGTGCTACAGGAACATGC	3000
Db	11896	TCCCATGTTGGAGACTCTAAGGATAAGGCTGGTGTGCCCGGAGGTGCTACAGGAACATGC	11955
Qy	3001	AGGTGTCTACCCCAAGTCTTCCCTCTCCCAAGCCAGGGTGGCAGAGGCACCTAGATC	3060
Db	11956	AGGTGTCTACCCCAAGTCTTCCCTCTCCCAAGCCAGGGTGGCAGAGGCACCTAGATC	12014
Qy	3061	CTGTGGAGTTTCAGAACCAACAAGCAACACACAGGCGCATAGTTGGCCTTTGGCCACTGC	3120
Db	12015	CTGTGGAGTTTCAGAACCAACAAGCAACACAGGCGCATAGTTGGCCTTTGGCCACTGC	12074
Qy	3121	CACCCACGGCCCTCCCTTTTGTGCTCCATGCTGGCATCTTCACTCCCTACCCCTTCCCCA	3180
Db	12075	CACCCACGGCCCTCCCTTTTGTGCTCCATGCTGGCATCTTCACTCCCTACCCCTTCCCCA	12134
Qy	3181	GCCACTGCTGCTCATTCAACTTCGTCCATGTCCTCCACTGTTCTTATCAGCAGGTGG	3240
Db	12135	GCCACTGCTGCTCATTCAACTTCGTCCATGTCCTCCACTGTTCTTATCAGCAGGTGG	12194
Qy	3241	CCCTGGGCATCAGAACACGCTCCCTCTGGCACAGGTGGCAGACACACTCAGAGCATGT	3300
Db	12195	CCCTGGGCATCAGAACACGCTCCCTCTGGCACAGGTGGCAGACACACTCAGAGCATGT	12254
Qy	3301	CTGGCTTTCTGGTGGTCCAGGCTCATCTGCTTCTGATTTCCTCCCTCCCGAGGCTCA	3360
Db	12255	CTGGCTTTCTGGTGGTCCAGGCTCATCTGCTTCTGATTTCCTCCCTCCCGAGGCTCA	12314
Qy	3361	TTTTCCCTTTTCTGCTACACATCCCTGCTACCTCCTCTCAGCCGTCACAGATTCT	3420
Db	12315	TTTTCCCTTTTCTGCTACACATCCCTGCTACCTCCTCTCAGCCGTCACAGATTCT	12374
Qy	3421	TCCTATCACAGGGATGCCAGTTCTATTGTGGG	3455
Db	12375	TCTATCACAGGGATGCCAGTTCTATTGTGGG	12409
RESULT 12	HSAC000384/c		
LOCUS	HSAC000384	123159 bp	DNA linear HTG 13-JUN-2002
DEFINITION	Homo sapiens chromosome 11 clone pDJ393o15, *** SEQUENCING IN PROGRESS ***, 8 unordered pieces.		
ACCESSION	AC000384		
VERSION	AC000384.2	GI:21406175	
KEYWORDS	HTG: HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 123159) Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Burbee, D., Davie, J., Davies, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S., Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.		
TITLE	Template		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 123159)		
AUTHORS	Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Burbee, D., Davie, J., Davies, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S., Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.		

RESULT 12
HSAC000384

LOCUS	HSNC000384	123159 bp	DNA	linear	HTG 13-JUN-2002
DEFINITION	Human sapiens chromosome 11 clone pD3393015, 8 unordered pieces.				
PROGRESS	***				*** SEQUENCING IN PROGRESS

ACCESSION AC000384
VERSION AC000384.2 GI:21406175
KEYWORDS HTG; HTGS_PHASE1.

SOURCE	ORGANISM
human.	Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	

REFERENCE
AUTHORS

Davie, J., Davies, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L., Garner, H.R., Gordon, M., Gotway, G., Grant, O., Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U., Neuber, I., Oldenburg, A.

NEWTON, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S., Schagamen, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.

TITLE

Template

JOURNAL
REFERENCE
AUTHORS
to 123159
Brady, P. A.
Evans, G. A.
Brigance, S.
Rumelhart, P.
Burke, D.

Braunstein, R., Bledsoe, J. F., Blingnac, S., Balmeister, K. R., Bulbee, D.,
Davies, J., Davies, C. J., Davis, C., English, C., London, T.,
Franklin, T. L., Garner, H. R., Gordon, M., Gotway, G., Grant, O.,
Hahner, L., Harris, J., Hinson, S., Megarity, C., Narayanaswamy, U.,
Reid, A.

Newton, J., O'Brien, K., Oliver, T., Patel, P., Probst, S., Rayner, S., Schageman, J., Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T. and Wilson, R.

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2730)
Strausberg, R.
Direct Submission
Submitted (04-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunarathne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 54 Row: j Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
 Location/Qualifiers
 1..2730
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5359019"
 /tissue_type="Eye, retina, mouse strain C57Bl/6"
 /clone_lib="NIH_MGC_94"
 /lab_host="DH10B"
 /notes="Vector: pCMV-SPORT6"
 <1..1272
 /codon_start=1
 /product="Unknown (protein for IMAGE:5359019)"
 /protein_id="AAH23077.1"
 /db_xref="GI:18606000"
 /translation="DLSLDSGLDSASCFOVSLMKSLANGRRVTICTIIHQPSAKLFEM
 FKLYILTSQCGIKFGKVNLPILPKGLGCPPELDPPIIEVASGEYGDLPMLFL
 RAVNGICLTMAEKSGKGNELPAHCTPPELDTIESHTFATSTIQFCILFRRTFL
 SILRDLVTLRHMVSHLVIGTILGLYLIHGDDASKVFNNTGELFFSMFLMEPALPM
 TVLTFFLENAVFRHNLVYTLKAYLTAMADVPQVVCVVCSIVVWMTGPQAPAE
 TSRELLFSALATATLVAOSLIGLIGCAASTSQVATFVGPVTAIPVLLESGFFVSFKT
 IPLYLQWSYLSVVRVYFGFLGTLITTYGMERGHUTCLDEQCFPRDPQIILRELDEVEAEK
 IMDFLQYVFIIFLLALVLYLVLRVKSER"
BASE COUNT 514 a 858 c 660 g 698 t
ORIGIN
Query Match 45.0%; Score 1554.8; DB 10; Length 2730;
Best Local Similarity 77.9%; Pred. No. 3e-308;
Matches 2157; Conservative 0; Mismatches 502; Indels 111; Gaps 19;
Qy 690 CAGTGGTCTGGATAGCGCCCTCTTGTTTCCAAAGTGGTGTCCTCATGAAGTCCCTGGGCACA 749
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 CACGGGTCTAGACAGCGCTCTTGTTTCCAAAGTGGTGTCCTCATGAAGTCCCTGGGCACA 74
Qy 750 GGGGGGGCGGTACCATCATCTGCAACATCCACAGCCAGTGCACAGCTCTTTTGAGATGTT 809
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 TGGGGGGCGGCACCGTCTATCTGCAACATCCACAGCCAGTGCACAGCTCTTTTGAGATGTT 134
Qy 810 TGACAAGCTCTACATCCTGAGCAGGAGTCACTGTCATCTTCAAAGAGAGTGCTACCAACCT 869
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 TGACAAGCTCTACATCCTGAGCCAGGGCAATGCATCTTCAAGGGCGTGGTTACCAACCT 194
Qy 870 GATCCCTCTATCAAAGGAGCTCGGCTTGGCATTTGCCCCACCTACGACCAACCCGCGTGACTT 929
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 GATTCCTATCTGAAGGGCTTGCGTTGCACTGTCCACCTACCAACACCCGCTGACTT 254

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:39:13 ; Search time 4538 Seconds
(without alignments)
12330.411 Million cell updates/sec

Title: US-10-072-621-4
Perfect score: 3455
Sequence: 1 gccaccatggcggagaagc.....atgcagttgtattgtggg 3455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	19.9	1856	11	BC030301 Mus muscu
2	634.8	18.4	851	14	BQ179452 UI-M-EW0-
3	603	17.5	778	13	BT131100 603351741
4	548	15.9	548	14	BQ636356 hd07h11.y
5	541.6	15.7	1045	14	BM925745 AGENCOURT
6	506.8	14.7	1428	11	AK018283 Mus muscu

7	463	13.4	820	12	BG298084
8	440.4	12.7	777	9	AU119946
9	406.6	11.8	974	12	BG292762
10	397.4	11.5	964	12	BE742274
11	383.2	11.1	885	14	BQ930068
12	376	10.9	879	12	BF159563
13	369	10.7	688	17	BQ6468
14	367.4	10.6	625	14	BQ769103
15	360.2	10.4	647	10	BE262142
16	352.4	10.2	747	12	BE315332
17	352	10.2	594	10	BE148782
18	345	10.0	810	13	BT136747
19	325	9.4	630	10	BB640447
20	313.6	9.1	695	14	BQ745197
21	313.6	9.1	720	10	BE260763
22	313	8.9	535	12	BE861999
23	305.4	8.8	716	12	BG747171
24	297.8	8.6	537	14	BQ637714
25	294.4	8.5	802	14	W21981
26	291.6	8.4	476	10	AW254434
27	291.6	8.4	482	9	AI712719
28	271.6	7.9	440	10	AW253072
29	264	7.6	426	9	AI711009
30	263.8	7.6	684	12	BG283492
31	262.6	7.6	422	9	AI712660
32	261.8	7.6	667	10	BB630594
33	258	7.5	567	9	AA944183
34	257.4	7.5	443	12	BF654958
35	255.8	7.4	554	12	BE948744
36	249.2	7.2	650	14	BQ640204
37	246.4	7.1	540	10	AW491055
38	242.8	7.0	400	9	AI716340
39	242.6	7.0	608	10	AV973769
40	242	7.0	601	12	BG083855
41	240.6	7.0	709	13	BM593317
42	238.6	6.9	371	14	F06912
43	234	6.8	304	14	BM692180
44	230.2	6.7	686	13	BM620534
45	229.6	6.6	500	14	BQ258653

ALIGNMENTS

RESULT 1
BC030301
LOCUS BC030301
DEFINITION Mus musculus, Similar to RIKEN cdNA 6430517004 gene, clone
IMAGE:5367539, mRNA.
ACCESSION BC030301.1 GI:20987866
VERSION BC030301
KEYWORDS HTC
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1856)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov
 Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.C., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgoun, C., Voigt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAP Plate: 65 Row: g Column: 11

This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis

This clone has the following problem: incomplete processing.

FEATURES source

Location/Qualifiers
 1. 1856
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5367539"
 /tissue_type="Eye, retina, mouse strain C57Bl/6"
 /clone_lib="NIH_MGC_94"
 /lab_host="DH10B"
 /note="vector: pCMV-SPORT6"
 BASE COUNT 386 a 545 c 472 g 453 t
 ORIGIN

Query Match 19.9%; Score 687; DB 11; Length 1856;

Best Local Similarity 70.6%; Pred. No. 7.7e-148;

Matches 1228; Conservative 0; Mismatches 400; Indels 111; Gaps 19;

```

QY 1721 GGTATGGCTTTGAGGGGTGATCTCGACGATCTATGGCATGGAGGAGGACCTGACAT 1780
DB 154 GGGATGTGCTTTGAGGGGTGATCTCGACCATCTATGTATGGAGCGAGACACCTGACCT 213
QY 1781 GTTTAGAGAACGCTGCTCCGGGAGCCAGACAGATCTCCGAGCGCTGGATGGG 1840
DB 214 GCTTAGATGACAGTGGCCATTCGGGAGCCACAAATCATTTTCGTGAGCTGGATGTAG 273
QY 1841 AGGATGCCAAGCTCTACATGACATCTCTGCTGCTTGGGATCTTCTTCTAGCCCTGGGGC 1900
DB 274 AGGAGGCCAAACTCTACATGACATCTCTGCTGCTTGGGATCTTCTTCTGCTCCCTGGGC 333
QY 1901 TGCTGGGCTACTTGTGCTGCTTACCGGGTCAAGTCAAGTCAAGATAGAGGCTTCCGCCAG 1960
DB 334 TGCTGGGATACCTTGTGCTGCTGCTTACCGGGTCAAGTCTGAGATAGAGGCTTGGCTCCTA 393
QY 1961 CCTGTACCCAGCCCTGACAGGAGAGCCGCCAGTCTCCAGCCCTTTGGGACTGTTTAA 2020
DB 394 CCTCTTC-----TGGSCAGAGAGGCCGCCAGTTCCAGCCCTTTGGGCTGTTTAA 443
QY 2021 CCTTATAGACTGGGCTGCTTCTGGCGGGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCT 2080
DB 444 ACTTGT---ACTTTGGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 501
QY 2081 CAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2137
DB 502 CAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561
QY 2138 CTCCCACTATGCTCCAGGAGTCTTCCCAAGTTGATGGGTTGTAGCTTCTCTCTCTCTCTCT 2197
DB 562 CTCCCGTC-RAGGCCAGGAGTCTTCCCAAGTTGATGGGTTGTAGCTTCTCTCTCTCTCTCT 620
QY 2198 TCTCCACACCTGATGCAAGACTACTGGGAGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 2257
DB 621 TCTCCACACCTGATGCAAGACTGCTGGGAGGTTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 679
QY 2258 CCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2317
DB 680 CCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 739

```

```

QY 2318 AAGTGGCCCTCTGGGGTCCCCACACACAAAGTGTGTTGTAACCTGGGCTGCTATAG 2377
DB 740 AGATTGACTCCCTCTGGGCTCCCGCCACAC-----TCTGCTATGATG 782
QY 2378 GTTGGAGTTCAGGGCTGGGCTGGTGGAGTCCACTGGAAGTCCCATTTAGTGGATGGA 2437
DB 783 GTTGGAGTTCAGGGCTGGGCTGGTGGAGTCCACTGGAAGTCCCATTTAGTGGATGGA 842
QY 2438 AATGACAGGAGGAGTCTGGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2497
DB 843 ATGGAGCAGGAGGAGTCTGGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 2498 ACCCTGGCTGACTTGGCAATCTCCAGGACAGAGTGGGCTTTCTCTCTAGTGGACCA 2557
DB 901 ACTTTGGCTAATTTGGACTATCTGTTGGACAGAGCGGGCTTTCTCTCTGGGCGCACCA 960
QY 2558 CTCCCAATCTCTGGGATTTGGAGAGGCTTGGGCTGTTGGATGTCCTCTCTCTCTCTCT 2617
DB 961 CTCCCAATCTCTGGAGATTGGAGAGCATAGGATGT-----CCCGGCCCTTCCCTA 1012
QY 2618 CACCTTTGGTGGGCGCAGGCGCTGGTGGCAGCTGTGCAATTAATGCTGTCTCTCTCT 2677
DB 1013 AGCTTTGGTCTGGATAGACCTCAACATCACTCTGCAATTAATGCTGTCTCTCTCTCT 1072
QY 2678 ACCTGCCACTGGAATGAGAAATGCACATTTATCTGGGCGGGGTGAGTGGGGAGAAC 2737
DB 1073 ATCGGCCACAGGACTGGAGAAATGCACATTTATCT-----GGGGGTGAGGCGGGA 1128
QY 2738 CCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2797
DB 1129 CCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1188
QY 2798 CCGAGTGCAGAGGACATACATAGAGACAGAGCCATCTCAGCCCTACACACTTGGCAT 2857
DB 1189 ATCATATCAGAGGACACAAAGCGCGGCAATCTCAGCCCTACACAGGACCATCTCC 1248
QY 2858 CTACAGGACAGAGGAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2917
DB 1249 TTGTGGCTCAGAGGAGTGGCGGTGGCA-----GGGAG 1283
QY 2918 CAGTGGCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2977
DB 1284 CAGTGGCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1342
QY 2978 CCAAGGTGTCTCAGAGAACTGCAGTGTCTACCCCAAGTCTTCCCTCTCTCTCTCTCT 3037
DB 1343 CCAAGTGTCTCAGAGAACTGCAGTGTCTACCCCAAGTCTTCCCTCTCTCTCTCTCTCT 1402
QY 3038 GGGTGGCAGAGGCACTAGATCCTCTGGAGTTCAGGAACCAACAGCAAGCAAGCAAG 3097
DB 1403 CGGG-----GTGGCGCTGGCGTTCGAGGTGAGAGTTCAGAGTTCAGAGTTCAGAG 1454
QY 3098 CATAGTTGGCTTTGGGCACTGCGCAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3157
DB 1455 CATAGTTGGCTTTGGGCACTGCGCAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1513
QY 3158 TTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3216
DB 1514 -TCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1572
QY 3217 TCCACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3276
DB 1573 TTCA-TCTTCGGTTCAGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1631
QY 3277 GTGCGAGACACTCAGAGCATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3336
DB 1632 GTGCGAGACACTCAGAGCATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1683
QY 3337 TGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3396
DB 1684 TGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1728
QY 3397 TCCTCTCACCCTGCCACAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3455

```

```

Db 1729 TCCTGTATCTCCCGAGTCTCTCTCATTTACACAGAGA--CCAGTGTATTGTGGG 1785
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 2
BQ179452
LOCUS
DEFINITION
  UI-M-EW0-bw-e-17-0-UI.r1 NIH_BMAP_EW0 Mus musculus cDNA clone
  IMAGE:5703928 5', mRNA sequence.
ACCESSION
  BQ179452
VERSION
  BQ179452.1 GI:20354957
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
REFERENCE
  1 (bases 1 to 851)
AUTHORS
  NIH-MGC http://mgc.nci.nih.gov/.
TITLE
  National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. James Lin, University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LNL at:
  http://image.llnl.gov
  This clone was contributed by the Brain Molecular Anatomy Project
  (BMAP)
FEATURES
  Seq primer: pYX-5.
  Location/Qualifiers
    1..851
      /organism="Mus musculus"
      /strain="C57BL/6"
      /db_xref="taxon:10090"
      /clone_lib="IMAGE:5703928"
      /clone_lib="NIH_BMAP_EW0"
      /tissue_type="whole brain"
      /dev_stage="embryo 15.5 dpc"
      /lab_host="DH10B (T1 phage resistant)"
      /note="Organ: Brain; Vector: pYX-Asc; Site:1: EcoR I;
      Site:2: Not I; The library was constructed according to
      Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
      1996. Denatured mRNA was size fractionated on a 1% agarose
      gel. First strand cDNA synthesis was primed with an
      oligo-dT primer containing a Not I site. Double stranded
      cDNA was size selected according to mRNA size fraction,
      ligated with EcoR I adaptor, digested with Not I, and then
      cloned directionally into pYX-Asc vector. The library tag
      sequence located between the Not I site and the polyA tail
      is GTGGCTGGAA. This library was created for the
      University of Iowa Mouse Brain Molecular Anatomy Project
      (BMAP). 'Gene Discovery in the Developing Mouse Nervous
      System', supported by National Institutes of Mental Health
      (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 168 a 264 c 205 g 210 t 4 others
ORIGIN
  Query Match 18.4%; Score 634.8; DB 14; Length 851;
  Best Local Similarity 88.9%; Pred. No. 5.9e-136;
  Matches 684; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
  QY 570 GGGCTGATGTCGTCTCCACAGGAGGAGCGCTCTCTCTGGCGGGGAGAGAGCG 629
  Db 1 GGGACTGATGTCGTCTCTCACAGGAGGAGCTCTCTCTCTGGGAGGAGAGAGCG 60
  QY 630 TCTGGCCATCGCTGGAGCTGTTCACAAACCGCGCTGTCTCTCTTTCATGAGCCAC 689
  Db 61 CTGGGCAATGTCGTCTGGAAGTGGTCAACACCGCGCTGTCTCTTTCATGAGCCAC 120
  QY 690 CAGTGTGCTGGATAGCGCCTCTTGTTCACAGTGGTGTCCCTCATGAAGTCCCTGGCACA 749

```

```

Db 121 CAGCGCTAGACAGCGCTTCTTGTTCCTCAAGTGGTGTCCCTCATGAAGTCCCTGGCACA 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 750 GGGGGCGGTACCATTATCTGCACCATTCACACAGAGCCAGTGCCTCTTTTTCAGATGTT 809
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TGGGGCGCCGACCGTCTATCTGCACCATTCACACAGCCAGTGCCTCTTTTTCAGATGTT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 810 TGACAAAGCTTACATCTCTGAGCCAGGCTCAGTGCATCTTCAAAGAGAGTGGTCAACCACT 869
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TGACAAAGCTTACATCTCTGAGCCAGGCGCAATGCATCTTCAAAGGGGCTGGTTACCACT 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 870 GATCCCTTCTAAAGGAGCTCGGCTTGCATTCGCCACCTACACAAACCCGCTGACTT 929
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GATTCCCTATCTGAAGGGGCTTGGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 930 CATCATCGAGTGGCTCTGCGCAGTATGAGACCTTGAACCCCATCTGTTCAGGCGCTGT 989
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CATCATTTGAGTGGCGCTCTGGAGAGTATGGAGACCTTGAACCCCATCTGTTCAGGCGCTGT 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 990 GCAGAAATGGGCTGTGGCTATGGCTGAGAAAGAGCAGCCCTGAGAAGAACAGAGTCCC 1049
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GCAGATGGGCTTTGCACCATGGCCGAGAGAGAGCAGCCCTGGGAAGATGAGCTCCC 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1050 TGGCCCATGCGCTCTTGTCTCGGAGTGGATGCCATTTGAAAGCCACACCTTTTGGCCAC 1109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TGGCCCATGCGCTCTTGTCTCGGAGTGGATGCCATTTGAAAGCCACACATTTTGGCCAC 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1110 CAGCACCCTCACACAGTCTTGCATCTCTTCAAGAGGAGCCTCTCTGCTCCATCTCAGGGA 1169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 CAGCACCCTAACCAAGTCTTGCATCTCTTCAAGAGGAGCCTCTCTGCTCCATCTCAGGGA 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1170 CAGCGTCTGACCCACCTACGCTTTCATGTCCACGCTGTTATTGGCGTGTCTCATCGGCT 1229
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 CAGCGTCTGACCCACCTTGCCTTTCATGTCCACATGTGTGATGGCGTCTCTCATTTGCGCT 660
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1230 CTCTACCTGCATATTGGCGAGCATGCCAAGGCTTTCACAAACACCGCGTGCCTCTT 1289
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CTCTACCTGCATATTGGCGAGTATGCCAGANAGTCTTCAACANACACCGCTTCTCTT 720
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1290 CTCTCCATCTGCTCTCATGTTCGCGCGCTCTCATGCTCCCACTGTGCTC 1338
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CTCTTCATCTCTTCTCTCATGTTTCAGCGCTCATNGCGAGTGTCTCAC 769
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 3
BQ1731100
LOCUS
DEFINITION
  BQ1731100.1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5359019 5',
  mRNA sequence.
ACCESSION
  BQ1731100
VERSION
  BQ1731100.1 GI:15708113
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
REFERENCE
  1 (bases 1 to 778)
AUTHORS
  NIH-MGC http://mgc.nci.nih.gov/.
TITLE
  National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: The Cepko Laboratory
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
  DNA sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LNL at:
  http://image.llnl.gov
  Plate: LLAM1913 row: b column: 12
  High quality sequence stop: 762.
  Location/Qualifiers
    1..778
FEATURES
  source

```

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5359019"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      157 a   249 c   179 g   193 t
ORIGIN
Query Match      17.5%; Score 603; DB 13; Length 778;
Best Local Similarity 89.1%; Pred. No. 1.3e-128;
Matches 67; Conservative 0; Mismatches 80; Indels 2; Gaps 2;
QY 690 CAGTGTCTGTGATAGCGCTCTTGTTCCTCAAGTGGTCTCCCTCATGAAGTCCCTGGCACA 749
DB 14 CAGCGGTCTAGACAGCGCTTCTTGTTCCTCAAGTGGTCTCCCTCATGAAGTCCCTGGCACA 73
750 GGGGGCGGTACCATCATCTGCACCATCCACAGCCAGTGCACCAAGCTCTTTGAGATGTT 809
DB 74 TGGGGCGCGCACCGTCTATCTGCACCATCCACAGCCAGTGCACCAAGCTCTTTGAGATGTT 133
810 TGACAAAGCTCTACATCTCTGAG-CCAGGTCAGTGCATCTTCAAGGAGTGTGCACCAACC 868
DB 134 TGACAAAGCTCTACATCTCTGAGACAGGGCAATGCATCTTCAAGGGGTGTGTACCAACC 193
869 TGATCCCTCTATTAAGGGGACTCGGCTTGCAATGGCCCCACCTACCAACACCCGGCTGACT 928
DB 194 TGATCCCTCTATTAAGGGGCTTGGCTTGCACTGTCCACCTACCAACACCCGGCTGACT 253
929 TCATCATGAGGTGGCTCTGGCGAGTATGGAGACCTGAACCCCATGTGTTTCAGGGCTG 988
DB 254 TCATCATGAGGTGGCTCTGGAGAGTATGGAGACCTGAACCCCATGTGTTTCAGGGCTG 313
989 TGCAGATGGGCTGTGCGCTTATGGCTGAGAAGAGCAGCCCTTGAGAAGCAGCGTCC 1048
DB 314 TGCAGATGGGCTTTCACCATGCGCGAGAGAGCAGCCCTTGAGAAGCAGCGTCC 373
1049 CTGCCCCATGCCCTCTTGTCTCGGAAGTGGATCCCATTTGAAGCCACACCTTTGCCA 1108
DB 374 CTGCCCCATGCCCTTGTGCTCGGAGCTGCGAGTGGATCTATTGAAGCCACACATTTGCCA 433
1109 CCAGCACCTTCACAGTCTGTGATCTCTTCAAGAGGACCTTCTGTCCATCTCAGGG 1168
DB 434 CCAGCACCTTCAACCCAGTCTGTGATCTCTTCAAGGAGGACTTCTGTCCATCTCAGGG 493
1169 ACAGGTCTGTGACCCACCTACGCTTCATGTCCACAGTGTATTGGCGTCTCATCGGC 1228
DB 494 ACAGGTCTGTGACCCACCTTGGGTTCTATGTACATGTGTGATCGGGTCTCATTTGGCC 553
1229 TCCTCTACCTGCATATTGGGAGCAGTGCAGCAGAGTCTTCAACACACCGGCTGCCTCT 1288
DB 554 TCCTCTACCTGCATATTGGGAGTATGCAGCAGAAAGTCTTCAACACACCGGCTTCTCT 613
1289 TCCTCTCCATGCTTCTTCATGTTTCGCGCCCTCATGCCAATGTGCTCACCTTCCCT 1348
DB 614 TCCTCTCCATGCTTCTTCATGTTTCGAGCCCTCATGCCGAGCGGTGCTCACCTTCCCT 673
1349 TAGAGATGGCGGTCTTTCATGAGGAGCAGCTCAACTACTGGTACAGCCTCAAGGCTATT 1408
DB 674 TAGAGATGGCGGTCTTTCATGAGGAGCAGCTTGTATCTACTGGTATATCTTAAAGCCTATT 733
1409 ACCTGCCAAGACCATGGCTGAGGTGCCCTTTCAG 1443
DB 734 A-CTTGGCAGACCATGGCTGATGTGCCCTTCCAG 767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hd07h11"
/clone_lib="Human Retina cDNA (Un-normalized, unamplified)"
/tissue_type="Retina"
/lab_host="EMDH10B"
/notes="Organ: Eye; Vector: pSPORT1; Neural retina tissue
eye dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library in the
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was constructed at
Bioserve Biotechnology (Laurel MD) essentially following
the protocols of the SuperScript Plasmid System full
details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor [5'-pGACTAGTTCAGTCCGAGCGGCCGCC(T)15-3'
]. EST analysis was performed on the unamplified library
at the NIH Intramural Sequencing Center (NISC)."
```

```

Db      241  |||||||ACTGAGGTGTACCCCAAGTCTTCCCTCCCAAGCCAGGGGTGGCAGAGGCACACT 300
QY      3056  AGATCCCTGAGGTTTCAAGAACCAACACAGCAACACAGGCGCATAAAGTTGGCCTTGGCC 3115
Db      301  AGATCCCTGAGGTTTCAAGAACCAACACAGCAACACAGGCGCATAAAGTTGGCCTTGGCC 360
QY      3116  ACTGCCACCCAGCGCCCTCTTTTGTGTCTCCATGCTGGCATCTTCACTCCCTTACCCCTT 3175
Db      361  ACTGCCACCCAGCGCCCTCTTTTGTGTCTCCATGCTGGCATCTTCACTCCCTTACCCCTT 420
QY      3176  CCCGAGCCACTGCTGCTATTAACACTTCTGTCCATGCTCCCTCCACTGTTCTCTATCAGCA 3235
Db      421  CCCGAGCCACTGCTGCTATTAACACTTCTGTCCATGCTCCCTCCACTGTTCTCTATCAGCA 480
QY      3236  GGTGGCCCTGGGCATCAGACAGCCTGCGCTGGGACAGGTGGCAGACACACTCAGAG 3295
Db      481  GGTGGCCCTGGGCATCAGACAGCCTGCGCTGGGACAGGTGGCAGACACACTCAGAG 540
3296  CATGCTG 3303
|||||||
541  CATGCTG 548

```

```

RESULT 5
BM925745
LOCUS
DEFINITION BM925745 1045 bp mRNA linear EST 12-MAR-2002
5' mRNA sequence.
AGENCOURT_6710400 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5763981

```

```

ACCESSION BM925745
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1045)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12816 row: k column: 22
High quality sequence start: 34
High quality sequence stop: 614.

```

```

FEATURES
Source
1..1045
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5763981"
/clone_lib="NIH_MGC_114"
/lab_host="PH10B"
/notes="Organ: brain; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dr
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invited). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."
BASE COUNT 185 a 334 c 367 g 159 t
ORIGIN
Query Match 15.7%; Score 541.6; DB 14; Length 1045;
Best Local Similarity 91.8%; Pred. No. 2.5e-114;

```

```

Matches 617; Conservative 0; Mismatches 49; Indels 6; Gaps 4;
QY      7  ATGGCGGAGAACGCGCTGGAGCCGCTGGGCTGTGGACTAGGCGCGGGGGCTGTGGCCATG 66
Db      347  ATGGCGGAGAACGCGCTGGAGCCGCTGGGCTGTGGACTAGGCGCGGGGGCTGTGGCCATG 406
QY      67  GCCGTGACGCTGGAGGAGCGGCGGAAACCCCTGTGCTGACACACACCTGAAAGAGTG 126
Db      407  GCCGTGACGCTGGAGGAGCGGCGGAAACCCCTGTGCTGACACACACCTGAAAGAGTG 466
QY      127  GAGAACACATCACTGAACCCAGCGCTTCTCCACCTGCCCAAGCGCTCAGCCGTGGAC 186
Db      467  GAGAACACATCACTGAACCCAGCGCTTCTCCACCTGCCCAAGCGCTCAGCCGTGGAC 526
QY      187  ATCGAGTTCGTGAGCTGTCTTATTCGCGGGAGGGGCGCTGTGGCGCAAAAGGGGT 246
Db      527  ATCGAGTTCGTGAGCTGTCTTATTCGCGGGAGGGGCGCTGTGGCGCAAAAGGGGT 586
QY      247  TATAAGACCCCTTCTCAAGTGCCTCTCAGTAAATTTGCGCGGGAGGAGCTGATTCGATC 306
Db      587  TATAAGACCCCTTCTCAAGTGCCTCTCAGTAAATTTGCGCGGGAGGAGCTGATTCGATC 646
QY      307  ATGGGCCCCCTCAGGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 366
Db      647  ATGGGCCCCCTCAGGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 706
QY      367  TCTGGAATGAAGGCGCAGATCTGTTTAAATGAAGGCCACGGAGCTGAGGACCTTCCGC 426
Db      707  TCTGGAATGAAGGCGCAGATCTGTTTAAATGAAGGCCACGGAGCTGAGGACCTTCCGC 766
QY      427  AAGATGTCTCTGCTACATCATGCAAGTACATGCTGCTGCCACACCTCACGGT-GTGGA 485
Db      767  AAGATGTCTCTGCTACATCATGCAAGTACATGCTGCTGCCACACCTCACGGTGTGGA 826
QY      486  ACCCATGATGCTCTCTGCTAACTGAATCTTACTGAGATCCCGATGTG-AAAAACCATC 544
Db      827  ACCCTGGAATGCTCTCTGCTAACTGAATCTTACTGAGATCCCGATGTG-AAAAACCATC 886
QY      545  TCGTGACACAGATCTCGACGGCCTGAGGCTGCTGCTGCCACACCTCACGGT-GTGGA 604
Db      887  TCGTGACACAGATCTCGACGGCCTGAGGCTGCTGCTGCCACACCTCACGGTGTGGA 946
QY      605  TGCTCTCTGCGGGG---CAGAGGAAGCGTCTGCCCATCCCTGGAGCTG-GTCAACAAC 660
Db      947  TGCTCTCTGCGGGGCAAGGGAAGCGTCTGCCCATCCCTGGAGGAGTGTGCAAAAAC 1006
QY      661  CCGCCTCTGTCATG 672
Db      1007  CCGCGGGGCAAG 1018

```

```

RESULT 6
AK018283
LOCUS
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430517004:hypothetical protein, full
insert sequence.
ACCESSION AK018283
VERSION AK018283.1 GI:12857918
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male olfactory brain cDNA to
mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
clone:6430517004.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2

```


307	ATGGGCCCCCTCAGGGCTGGCAAGCTACATTCATGAACTCTTGGCAGGATACAGGAG	366
Qy		
420	ATGGGCCCCCTCGGGCTGGCAAGCTACACATTCATGAACATCTTGGCAGGGTACAGGAG	479
Db		
367	TCGTGAATGAAGGGGCAGATCCTGGTTAATGGAAGGCCACGGGAGCTCAGAGACCTTCGC	426
Qy		
480	TCGGGGATGAAGGGGCAGATCCTGGTTAACGGGAGGCCGCGAGCTCAGAGACCTTCGC	539
Db		
427	AAGATGTCCTGTACATCATGCAAGATGACATGCTGTGTG--CCGCACCTCACGGTGTGG	484
Qy		
540	AAGATGTCCTGTACATCATGCAAGATGACATGCTGTGTGTCTCAACCTCAGTCCCTGG	599
Db		
485	-AGCCATGATGTCCTGTCTAACTGAATCTTACTGAGAAATCCGGATGTGAAACAGAT	543
Qy		
600	CAGGCCATGATGTGTCTCTGCGCAACTGAAAGCTGAGTGAGAGCAAGAGGTGAAGAAGAA	659
Db		
544	CTGCTGACAGAGATCCGTG--ACGGCACTGGGCCCTGATGCTGTGTCTCCACACAGG-ACA	600
Qy		
660	CTGCTGACAGAGATCCGTGAGACCCCTGGGACTTGTATGTCCTCTCACACAGGAGACA	719
Db		
601	GCCCTGCTCTCTGCGGGCGAGGAAGCCTCTGGCCATCGGCCCTGGAGCTGGTCAACAC	660
Qy		
720	GATCTGCTCTCTGGAGGCGAAGCGACGCCCTGGCATTGGTCTTCTGGGACTGTTTACACA	779
Db		
661	CCGCCTCTCATGTTCTTTTGTATGAGCCCAACAGTGG	695
Qy		
780	ACCGCCTGTATGTCTTTTGTATGAACCTACCAACGG	814
Db		

RESULT 8	AU119946	777 bp	linear	EST 01-AUG-2002
AU119946				
LOCUS	AU119946	777 bp	mRNA	
DEFINITION	HEMBAL Homo sapiens cDNA clone HEMBAL1006999 5', mRNA			
ACCESSION	AU119946			
VERSION	AU119946.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			

[illegible][illegible]

cdna Library Preparation: Life Technologies, Inc.
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10368 row: j column: 12
 High quality sequence stop: 627.
 High quality sequence stop: 627.

FEATURES

Location/Qualifiers
 1..974
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4501355"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="organ: eye; Vector: PCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 199 a 312 c. 253 g 210 t

BASE COUNT

Query Match 11.8%; Score 406.6; DB 12; Length 974;
 Best Local Similarity 84.7%; Pred. No. 4.1e-83;
 Matches 526; Conservative 0; Mismatches 89; Indels 6; Gaps 6;

QY 690 CAGTGGCTGGATAGCGCTCTGTGTTCCAAAGTGTGTCCTCATGAAGTCCCTGGCACA 749
 DB 140 CAGCGGTCTAGACAGCGCTCTGTGTTCCAAAGTGTGTCCTCATGAAGTCCCTGGCACA 199
 QY 750 GGGGGGCCCTACCATCATCTGCACCATCCACCAGCCAGTGCACCAAGCTCTTTGAGATGTT 809
 DB 200 TGGGGGCCCGACCGTCATCTGCACCATCCACCAGCCAGTGCACCAAGCTCTTTGAGATGTT 259
 QY 810 TGCAAGCTCTACATCTGAGCCAGGTCAGTGCATCTTCAAGAGTGGTGCACCAACT 869
 DB 260 TGCAAGCTCTACATCTGAGCCAGGTCAGTGCATCTTCAAGAGTGGTGCACCAACT 318
 QY 870 GATCCCTCATCTAAAGGACTCGGCTGCTGATTTGCCCCACCTACCACACCCCGCTGACTT 929
 DB 319 GATTCCTCATCTGAAGGGGCTTTGGCTTGCACTGTCCCACTACCACACCCCGCTGACTT 378
 QY 930 CATCATGAGTGGCTCTGGCGAGTATGGAGACCTGAACCCCATGTTCTTCAGGCTCT 989
 DB 379 CATCATGAGTGGCTCTGGAGAGTATGGAGACCTGAACCCCATGTTCTTCAGGCTCT 438
 QY 990 CGAAGTGGCTGTGGCTATGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
 DB 439 CGAAGT-GGCTTTGACCATGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 497
 QY 1050 TCGCCCATGCGCTCTGCTTCGCGAGTGGATCC-CATTGAAGCCACACTTTGCA 1108
 DB 498 TCGCCCATGCGCTCTGCTTCGCGAGTGGATCC-TAATGAAAGCCACACTTTGCA 557
 QY 1109 CCAGCACCCTACACAGTCTTCATCTCTTCAAGAGGAGCTTCCTGTCCATCTCCAGG 1168
 DB 558 CCAGCA-CCTAACCAGTCTTCATCTCTTCAAGAGGAGCTTCCTGTCCATCTCCAGG 616
 QY 1169 ACAGGCTCTGACCCACCTACGGTTCATGTCACAGTGGTATATGGCGTGTCTATGGGCC 1228
 DB 617 ACAGGCTCTGACCCACTTGGCGTTCATGTCACATGT-CTGATCGGCGTCTCAT-GGCT 674
 QY 1229 TCTCTACTGTCATATGGCGAGGATGCCAGCAAGCTCTTCAACACACCGGCTGCTCT 1288
 DB 675 CCCTCTACTGGACATAGGAGGATGAGGCCAGGCAAGTCTTCAACACACCGGCTGCTCT 734
 QY 1289 TCTCTCTACTGCTGTTCCTCA 1309
 DB 735 TCTTTTCCAGGCTCTCTCTCA 755

RESULT 10

LOCUS BE742274

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE742274
 601575692F1 NIH_MGC_9 Homo sapiens cdna clone IMAGE:3836615 5',
 mRNA sequence.
 BE742274
 BE742274.1 GI:10156266
 EST.
 human.
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 964)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: DCTD/DTP
 cdna Library Preparation: Ling Hong/Rubin Laboratory
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
 Plate: LLCM520 row: h column: 24
 High quality sequence stop: 681.

FEATURES

Location/Qualifiers
 1..964
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3836615"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="organ: ovary; Vector: pORF7; Site_1: XhoI; Site_2:
 EcoRI; cdna made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(C). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 181 a 303 c 277 g 203 t

Query Match 11.5%; Score 397.4; DB 12; Length 964;
 Best Local Similarity 78.0%; Pred. No. 5.5e-81;
 Matches 492; Conservative 0; Mismatches 136; Indels 3; Gaps 1;
 QY 1234 TACCTGTCATTTGGCGACGATGCCAGAGTCTTCAACAACACCGCTCTCTCTCTTC 1293
 DB 1 TACTTGGGATCGGGNACGAGCCAGAGAGGCTTTGAGCAACTCCGGCTCTCTCTCTTC 60
 QY 1294 TCCATGCTCTTCTCTCATGTTTCGCGCCCTCATGCCCACCTGCTCACTACCTTCCCGCTTAGAG 1353
 DB 61 TCCATGCTCTTCTCTCATGTTTCGCGCCCTCATGCCCACCTGCTCACTACCTTCCCGCTTAGAG 120
 QY 1354 ATGGCGGCTCTTCTGAGGAGCAGCTCACTACTGTCAGCCTCAAGCGTATTACCTG 1413
 DB 121 ATGGGAGTCTTCTTCTGGGAACACCTGAACACTACTGTCAGCCTGAGGCGCTACTACCTG 180
 QY 1414 GCCAAGACCATGGCTGACGTGCCCTTTTCAGGTGGTGTGTCGGGTGCTTACTGACGACTT 1473
 DB 181 GCCAAGACCATGGCAGACGCTGCCCTTTTCAGATCATGTTCCCGAGTGGCTTACTGACGATC 240
 QY 1474 GTGTACTGATGACGGGCGACCGCTGAGACAGCGCTTCTGCTCTTCTTCTGACGCTG 1533
 DB 241 GTGTACTGATGACGTCGACGCGCTCCGACGCGCTGCTGCTGCTGCTGCTGCTGCTG 300
 QY 1534 GCCACCGCCACCGCTTGGTGGCCCACTTTTGGGCTCTGCTGATGGAGCTGCTTCCAC 1593
 DB 301 GGCACCATGACCTCCCTGCTGGTGCACAGTCCCTGGGCTCTGCTGATGGAGCCCGCTCCAG 360

QY 20

JOURNAL
COMMENT

NIH-MGC HCLP
National Ins
Unpublished

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 879)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

604	DB	CAGCGCCCTCCGCGCTCCACAGTGGTGTCTGTATGAGAGGACTGGCCCGAGGGGTGGCGCT	663
761	QY	CCATCATCTGCACCATCCACCAGCCCACTCTTTAGAGTCTTTGAGATGTTTGACAAGCTCT	820
664	DB		722
821	QY	ACATCGCTCTGTACCATCCACACAGCAGTCTTTAGAGCTCTTTGACCAGCT-T	880
723	DB		782
881	QY	TAAGGGACTCGGCTTGCAATGCC	904
783	DB	TGAGGGACCGGGTCTGAAATGCC	806

Qy

QY Db

SOURCE
ORGANISM

REFERENCE

AUTHORS

TITLE

**JOURNAL
COMMENT**

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Ma

Best Local Matches

Oy

7 2

25

QY

66	DB	CACTGCCACCGGCCCTCTTTGTGCTCCATGCTGGCATCTGACTCCCCCTACCCCT	125
3175	QY	TCCCCAGCCACTGCTGCTCATTTCAAACATTCGTGCTCCATGTCCTCCACTGTTCTCTATCAGC	3234
126	DB	TCCCCAGCCACTGCTGCTCATTTCAAACATTCGTGCTCCATGTCCTCCACTGTTCTCTATCAAC	185
3235	QY	AGGTGGCCCTGGGCATCAGAACAGCCTGCCCTGGGCACCAGGTGGCAGCACACATCAGA	3294
186	DB	AGGTGGCCCTGGGCATCAGAACAGCCTGCCCTGGGCACCAGGTGGCAGCACACATCAAA	245
3295	QY	GCATGCTGGCTTTCCTGGTGCGTCCAGGCTCATTTGCTTCTTGATTTCCCTCCGCCAG	3354
246	DB	ACATGCTGGCTTTCCTGGTGCGTCCAGGCTCATTTGCTTCTTGATTTCCCTCCGCCAG	305
3355	QY	GGCTCATTTTCCCTTTTTCCTGTGACACATCCCTGTCTACCTCCCTCACCCTGGCACA	3414
306	DB	GGCTCATTTTCCCTTTTTCCTGTGACACATCCCTGTCTACCTCCCTCACCCTGGCACA	365
3415	QY	GATTTCTCCTATCACACAGGGATGCCAGTTGTAATTTGTGGG	3455
366	DB	GATTTCTCCTATCACACAGGGATGCCAGTTGTAATTTGTGGG	406

RESULT 14

BQ769103	BQ769103	625 bp	mRNA	linear	EST 26-JUL-2002
LOCUS	UI-M-FC0-byn-f-19-0-UI.r1		NTH-BWAP_FC0	Mus musculus	cdna clone
DEFINITION	IMAGE:5720466 5', mRNA sequence.				

ACCESSION BQ769103
VERSION BQ769103.1 GI:21977577
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 625)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

The following repetitive elements were found in this cDNA sequence: 2-55, >MTD#LTR/MaLR (matched complement)
Seq primer: NVX-5

```

FEATURES
source
Location/Qualifiers
1. .625
    /organism="Mus musculus"
    /strain="C57BL/6"
    /db_xref="taxon:10090"
    /clone="IMAGE:5720466"
    /clone_lib="NIH_BMAP_FC0"
    /tissue_type="whole brain"
    /dev_stage="embryo 12.5 dpc"
    /lab_host="DH10B (T1 phage resistant)"
    /note="Organ: brain; Vector: pYX-Asc; Site_1: Ecor I;
    Site_2: Not I; The library was constructed according to
    Bonatido, Lennon and Soares, Genome Research, 6:791-806,
    1996. Denatured mRNA was size fractionated on a 1% agarose
    gel. First strand cDNA synthesis was primed with an
    oligo-dT primer containing a Not I site. Double stranded
    cDNA was size selected according to mRNA size fraction,
    ligated with Ecor I adaptor, digested with Not I, and then
    cloned directionally into pYX-Asc vector. The library tag
    sequence located between the Not I site and the polyA tail

```

BASE COUNT 127 a 207 c 142 g 145 t
, is TGAGAGACC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BWAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT	127 a	207 c	142 g	145 t	4 others
ORIGIN					

Query Match	10.6%	Score 367.4;	DB 14;	Length 625;
Best Local Similarity	89.8%;	Pred. No. 3.8e-74;		
Matches 403.	Conservative	0.	Mismatches 45.	Indels 1.

	1097	CONSERVATIVE	G	MISMATCHES	49	INCRS	1	SUPS	1
QY	690	CAGTGGTCGGATAGCGCCTCTGTTTCCAAGTGGTGCCCTCATGAAGTCCTCCGGCAC							749

Db 178 CAGCGGTCAGACAGCGCTTCTTTGTTTCAAGTGGTGCTCCATGAAGTCCCTGGCCCCA 237

[illegible]

Qy 810 TGCAAGCTCTACATCCTGAGCCAGGTCAGTGCATCTTCAAGGAGTGGTCACCAACCT 869

Db	298	GGACAAGCTCTACATCCTGAGCCAGGCGCAATGCATCTTCAAGGGCGTGTTACCAACCT	357
Qy	870	GATCCCCCTATCTAAAGGGACTTCGGCTTGCAITGCCCCACCTACCAACAACCCGGTGACTT	929

Db
358 GATCCCTATCTNAGGGGGCTTGGCTTGCACTGTCACCTACCAACACCCCGCTNACTT 417

QY	930	CATCATCGAGGTGGCCTCTCGCGAGTATGGAGACCTGAACCCCATGTTCTTCAGGGCTGT	989
Dδ	418	CATCATTTNAGGTGGCCTCTCGAGAGTATGGAGACCTGAACCCCATGTTCTTCAGGGCTGT	477

Qy 990 GCAGAATGGGCTCTGCGCTATGGTGAGAGAAGAGCAGCCCTGAGAGAACGAGGTCCC 1049

478 GCAGAAATGGCCTTTTGACACATGGCCGAGAGGAAGAGCAGCCCTGGGAGAAATGAGCTCC 537

Db

1050 TGGCCCCATGGCCCCCTCTCTCTGTCCTCCGCAAGCTGCATCCCATTTGAAAGCCACACACCTTTTGGCCAC 1109

OV

Db	538	TGCCCCACTGCCCACTTGGCCCTCGGAGCTGGATCCTATTGAAAG-CACACATTTGCCAC	596
Qy	1110	CAGCACCCCTCACACAGTTCTGCACTCCTCT	1138
Db	597	CAGCACCCCTNAGCCAGTTCTGCATCCTCT	625

RESULT 15

BE262142	BE262142	647 bp	linear	EST 26-OCT-2000
LOCUS	60115187F1			
DEFINITION	NIH_MCC_19 Homo sapiens cDNA clone IMAGE:3508053 5', mRNA sequence.			

ACCESSION	BE262142
VERSION	BE262142.1
KEYWORDS	GI:9134906
EST.	
SOURCE	human.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 647)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D.

Email: cgabps.r@mail.nih.gov
Tissue procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC1M88 row: n column: 22
High quality sequence stop: 644.
Location/Qualifiers
1. .647

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:59:44 ; Search time 478 Seconds
(without alignments)
10466.632 Million cell updates/sec

```

Title:
Perfect score: 3455
Sequence: 1 gccacatggcggagaaagc.....atgccagttgtatttattggg 3455

```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

rchd: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA**

```

1: /cgn2_5/ptodata/2/pubpna/us07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCr_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/us06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/us06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/us07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/us08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/us08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/us09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/us09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/us10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/us10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/us60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/us60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	3455	100.0		3455	9	US-10-072-621-4	Sequence 4, Appli
2	3455	100.0		3455	9	US-10-090-455-1	Sequence 1, Appli
3	2592.2	75.0		2687	9	US-10-154-452-3	Sequence 3, Appli
4	2591.4	75.0		2687	9	US-10-090-455-12	Sequence 12, Appli
5	2589	74.9		2687	9	US-10-154-452-7	Sequence 7, Appli
6	1941	56.2		1941	9	US-10-090-455-3	Sequence 3, Appli
7	1062.4	30.7		3376	9	US-10-037-370-918	Sequence 918, App
8	972.8	28.2		2930	9	US-09-954-531-591	Sequence 591, App
9	972.8	28.2		2930	9	US-10-171-581-276	Sequence 276, App
10	972.8	28.2		3201	9	US-10-072-621-5	Sequence 5, Appli
c 11	238	6.9		287	10	US-09-864-761-28360	Sequence 28360, A
c 12	217	6.3		491	10	US-09-864-761-11789	Sequence 11789, A
13	107.2	3.1		2340	9	US-09-989-981A-5	Sequence 5, Appli
14	107.2	3.1		2340	10	US-09-837-993-4	Sequence 4, Appli
15	105.8	3.1		2400	9	US-10-108-605-244	Sequence 244, Appli
16	104.2	3.0		10330	9	US-10-001-189-68	Sequence 68, Appli
17	92.8	2.7		2025	10	US-09-866-866A-13	Sequence 13, Appli
18	87.6	2.5		2019	9	US-09-989-981A-3	Sequence 3, Appli
19	87.2	2.5		2669	9	US-09-989-981A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-10-072-621-4

Sequence 4, Application US/10072621

; Patent No. US20020169137A1

; GENERAL INFORMATION:

APPLICANT: Reiner, Peter B.

APPLICANT: Connop, Bruce P.

APPLICANT: Pollard, Michelle

; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION.

; TITLE OF INVENTION: BY MODIFICATION OF A
SEE REFERENCE 1,001,000-1,000,000

; FILE REFERENCE: 100103.402

; CURRENT APPLICATION NU

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846

; NUMBER OF SEQ

; SOFTWARE: Fas

```
; SEQ ID NO 4
; LENGTH: 3455
```

```

; LENGTH: 3455
; TYPE: DNA

```

; TYPE: DNA
; ORGANS: H

Query Match	100.0%;	Score 3455;	DB 9;	Length 3455;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3455; Conservative	0;	Mismatches	0;	Indels
	0;	Gaps	0;	
QY	1	GCACCATGCGGAGAAAGCGCTGGAGGCGCTGGGCTGTGGACTAGGGCGGGGCTGTG	60	
DB	1	GCACCATGCGGAGAAAGCGCTGGAGGCGCTGGGCTGTGGACTAGGGCGGGGCTGTG	60	
QY	61	GCATGCGCGTGACCGTGGAGGAGCGGGGGGGAACCCCTGTGTGTGACACGCACCTGAAG	120	
DB	61	GCATGCGCGTGACCGTGGAGGAGCGGGGGGGAACCCCTGTGTGTGACACGCACCTGAAG	120	
QY	121	AAGGTGGGAACACATCACTGAAGCCAGCGCTTCTCCACCTGCCCAAGCGCTCAGCC	180	
DB	121	AAGGTGGGAACACATCACTGAAGCCAGCGCTTCTCCACCTGCCCAAGCGCTCAGCC	180	
QY	181	GTGGACATCGAGTTCGTGGAGCTGCTTATTCGTGCGGAGGGGCCCTGCTGGCGCAA	240	
DB	181	GTGGACATCGAGTTCGTGGAGCTGCTTATTCGTGCGGAGGGGCCCTGCTGGCGCAA	240	
QY	241	AGGGGTATTAAGACCCCTTCTCAAGTGCCTCTCTCAGGTAAATTCGTCCGCGGGAGCTGAT	300	

|||||
781 CAGCCAGTGCCAAAGCTCTTTGAGATGTTTGACAAAGCTCTACATCCTGAGCCAGGGTCAg 840
Db
|||||
841 TGCATCTTCAAGAGGAGTGTGTCACCAACCTGATCCCTATCTAAAGGAGCTCGGCTTGCAAT 900
QY
|||||
841 TGCATCTTCAAGAGGAGTGTGTCACCAACCTGATCCCTATCTAAAGGAGCTCGGCTTGCAAT 900
Db
|||||
901 TGCCCACTACCAACACCGGCTGACTTTCATCATCGAGGTGGCTCTGCGCAGATGGA 960
QY
|||||
901 TGCCCACTACCAACACCGGCTGACTTTCATCATCGAGGTGGCTCTGCGCAGATGGA 960
Db
|||||
961 GACCTGAACCCCATGTTGTCAGGGCTGTGCAAGATGGGCTGTGCGCTATGGCTGAGAG 1020
QY
|||||
961 GACCTGAACCCCATGTTGTCAGGGCTGTGCAAGATGGGCTGTGCGCTATGGCTGAGAG 1020
Db
|||||
1021 AAGAGCAGCCCTGAGAGAACAGGTGCTTCCCTGCCCCATGCCCCCTTGTCTCCGGAAGTG 1080
QY
|||||
1021 AAGAGCAGCCCTGAGAGAACAGGTGCTTCCCTGCCCCATGCCCCCTTGTCTCCGGAAGTG 1080
Db
|||||
1081 GATCCATTTGAAGCCACACCTTTGCCCACAGCACCTTCACACAGTTCTGCAATCTCTTTC 1140
QY
|||||
1081 GATCCATTTGAAGCCACACCTTTGCCCACAGCACCTTCACACAGTTCTGCAATCTCTTTC 1140
Db
|||||
1141 AAGAGACCTTCTCTGTCATCTCAGGGACACGGTCTGACCCACCTACGGTTTCATGTCC 1200
QY
|||||
1141 AAGAGACCTTCTCTGTCATCTCAGGGACACGGTCTGACCCACCTACGGTTTCATGTCC 1200
Db
|||||
1201 CAGGTGTTATTTGGGCTGCTCATCGGCTCTCTACCTGCGATATTGGGACGATGCCAGC 1260
QY
|||||
1201 CAGGTGTTATTTGGGCTGCTCATCGGCTCTCTACCTGCGATATTGGGACGATGCCAGC 1260
Db
|||||
1261 AAGGTCTTCAACAAACACGGCTGCTCTCTCTCATGCTGCTCATGTTTCGCGCC 1320
QY
|||||
1261 AAGGTCTTCAACAAACACGGCTGCTCTCTCTCATGCTGCTCATGTTTCGCGCC 1320
Db
|||||
1321 CTCATGCCAATCTGCTCACCCTTCCCTTTAGAGATGGCGGTCTTCATGAGGAGCACCTC 1380
QY
|||||
1321 CTCATGCCAATCTGCTCACCCTTCCCTTTAGAGATGGCGGTCTTCATGAGGAGCACCTC 1380
Db
|||||
1381 AACTACTGTACAGCTCAAAAGCTATTACCTGGCCAGACCATGGCTGACCTGCCCCCTT 1440
QY
|||||
1381 AACTACTGTACAGCTCAAAAGCTATTACCTGGCCAGACCATGGCTGACCTGCCCCCTT 1440
Db
|||||
1441 CAGGTGTTGTCGCGGTGCTTACTGACGATTTGTACTGGATGACGGCCAGCCGCT 1500
QY
|||||
1441 CAGGTGTTGTCGCGGTGCTTACTGACGATTTGTACTGGATGACGGCCAGCCGCT 1500
Db
|||||
1501 GAGACAGCCGCTTCCCTCTTCTCAGGCTGGCCAGCCGCTTGGTGGCCCAA 1560
QY
|||||
1501 GAGACAGCCGCTTCCCTCTTCTCAGGCTGGCCAGCCGCTTGGTGGCCCAA 1560
Db
|||||
1561 TCTTTGGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGGTGGCCACCTTTCTGGGC 1620
QY
|||||
1561 TCTTTGGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGGTGGCCACCTTTCTGGGC 1620
Db
|||||
1621 CCAGTTACCGCCATCCCTGCTCTGTTGTCGCGCTTCTTGTGAGCTTCAAGACCATC 1680
QY
|||||
1621 CCAGTTACCGCCATCCCTGCTCTGTTGTCGCGCTTCTTGTGAGCTTCAAGACCATC 1680
Db
|||||
1681 CCCACTTACCTGCAATGAGCTCTATCTCTCCTATGTCAGATGAGCTTGGAGGTGTG 1740
QY
|||||
1681 CCCACTTACCTGCAATGAGCTCTATCTCTCCTATGTCAGATGAGCTTGGAGGTGTG 1740
Db
|||||
1741 ATCTGACGATCTATGGCATGGAGGAGGACCTGACATGTTTATAGAGAACGCTGCCCG 1800
QY
|||||
1741 ATCTGACGATCTATGGCATGGAGGAGGACCTGACATGTTTATAGAGAACGCTGCCCG 1800
Db
|||||
1801 TTCGGGAGCCACAGAGCATCTCCGAGCGTGGATGTGGAGATGCCAAGCTCTACATG 1860
QY
|||||
1801 TTCGGGAGCCACAGAGCATCTCCGAGCGTGGATGTGGAGATGCCAAGCTCTACATG 1860
Db
|||||
1861 GACTTCTTGGTGGGCTCTTCTAGCCCTGCGGCTGCTGGCCCTACCTTGTGCTG 1920
QY
|||||

1861 GACTTCTTGGGCTCTTGGGCTCTTCTTCTAGCCCTGCGGCTGCTGGCCCTACCTTGTGCTG 1920
Db
|||||
1921 CATTACCCGGGTCAAGTCAAGAGATAGAGGCTTGGCCAGGCTGTACCCAGCCCTTGA 1980
QY
|||||
1921 CATTACCCGGGTCAAGTCAAGAGATAGAGGCTTGGCCAGGCTGTACCCAGCCCTTGA 1980
Db
|||||
1981 GCAGGAAGCCCAAGTCCAGCCCTTTGGGACTGTTTAACTTATAGACTTGGGACATG 2040
QY
|||||
1981 GCAGGAAGCCCAAGTCCAGCCCTTTGGGACTGTTTAACTTATAGACTTGGGACATG 2040
Db
|||||
2041 GTTCTTGGGGGCTATCT 2100
QY
|||||
2041 GTTCTTGGGGGCTATCT 2100
Db
|||||
2101 GCTTCCAGGCTGGGCTCTGGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT 2160
QY
|||||
2101 GCTTCCAGGCTGGGCTCTGGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT 2160
Db
|||||
2161 TCCCAAGTTGATGCGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
QY
|||||
2161 TCCCAAGTTGATGCGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
Db
|||||
2221 CTACTGGGAGGCTGCTGCT 2280
QY
|||||
2221 CTACTGGGAGGCTGCTGCT 2280
Db
|||||
2281 AGCCCTAGGCTCTTAGGGCCCACTTACAACTGACCAAGTGGCCCTCTGGGGGCTCC 2340
QY
|||||
2281 AGCCCTAGGCTCTTAGGGCCCACTTACAACTGACCAAGTGGCCCTCTGGGGGCTCC 2340
Db
|||||
2341 CCACCACACAAGTGTGTTAACTGGGCTGCTTAAAGTTGGAGTTTCAGGGCTGGGCTCC 2400
QY
|||||
2341 CCACCACACAAGTGTGTTAACTGGGCTGCTTAAAGTTGGAGTTTCAGGGCTGGGCTCC 2400
Db
|||||
2401 TGTGGAGTCCACTGGAAGTCCCAATTTAGATGTTGAAATGGACAGGAAAGGACTCTGGA 2460
QY
|||||
2401 TGTGGAGTCCACTGGAAGTCCCAATTTAGATGTTGAAATGGACAGGAAAGGACTCTGGA 2460
Db
|||||
2461 AGTCTTCTTCT 2520
QY
|||||
2461 AGTCTTCTTCT 2520
Db
|||||
2521 GCCAGCAGAGCTGGGTTTTCTGTCTAGGTGACCACTTCCCAATCTCTGGGATTTGGAGA 2580
QY
|||||
2521 GCCAGCAGAGCTGGGTTTTCTGTCTAGGTGACCACTTCCCAATCTCTGGGATTTGGAGA 2580
Db
|||||
2581 GGCCTGGGCTGTGGGATGCCCATCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCC 2640
QY
|||||
2581 GGCCTGGGCTGTGGGATGCCCATCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCCCTCTCC 2640
Db
|||||
2641 GGTGGCACTGTGCAATTAATGCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2700
QY
|||||
2641 GGTGGCACTGTGCAATTAATGCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2700
Db
|||||
2701 GCATCTTATTTCTGGGCGGGGTGAGTGGGGAAGACCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
QY
|||||
2701 GCATCTTATTTCTGGGCGGGGTGAGTGGGGAAGACCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
Db
|||||
2761 CTAGCGCATGACGGTCTCTGTTGATGCT 2820
QY
|||||
2761 CTAGCGCATGACGGTCTCTGTTGATGCT 2820
Db
|||||
2821 AGACAGCCCATCTCAGCCCTACACCTGCT 2880
QY
|||||
2821 AGACAGCCCATCTCAGCCCTACACCTGCT 2880
Db
|||||
2881 GGTGCATGCTGGTGGGCTGCTGTTGGGAGGACAGTCCCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
QY
|||||
2881 GGTGCATGCTGGTGGGCTGCTGTTGGGAGGACAGTCCCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 2940
Db
|||||
2941 TCCCATGTTGAGACTCTAAGGATTAAGGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
QY
|||||
2941 TCCCATGTTGAGACTCTAAGGATTAAGGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
Db

QY	421	TTCCGCAAGATGTCCTGCTACATCATGCAAGATGACATGCTGCTGCCGCACTTCACGGTG	480
DB	465	TTCCGCAAGATGTCCTGCTACATCATGCAAGATGACATGCTGCTGCCGCACTTCACGGTG	524
	481	TTGGAAGCCATGATGGTCTCTGCTAACCTGAATCTTACTGAGAAATCCCGATGCAAAAC	540
	525	TTGGAAGCCATGATGGTCTCTGCTAACCTGAAGCTGAGTGAGAAGCAGGAGGTGAAGAAG	584
QY	541	GATCTGCTGACAGAGATCCTGAGGGCACTGGGGCTGATGTCTGCTCCACACAGAGACA	600
DB	585	GAGCTGCTGACAGAGATCCTGACGGCACTGGGGCTGATGTCTGCTCCCGCAGAGACA	644
	601	GCCCTGCTCTTGGCGGGCAGAGAAGCGTCTGGCCATCGCCCTGGAGCTGAGTCAACAAC	660
	645	GCCCTGCTCTTGGCGGGCAGAGAAGCGTCTGGCCATCGCCCTGGAGCTGAGTCAACAAC	704
QY	661	CCGGCTGCTCATGTTCTTGTGATGAGCCACACAGTGGTCTGATAGCGCCTCTTCTTCCAA	720
DB	705	CCGGCTGCTCATGTTCTTGTGATGAGCCACACAGTGGTCTGATAGCGCCTCTTGTTCCAA	764
	721	GTGGTGTCCCTCATGAAGTCCCTGGCACAGGGGGCGGTACCATCATCTGCACCATCCAC	780
	765	GTGGTGTCCCTCATGAAGTCCCTGGCACAGGGGGCGGTACCATCATCTGCACCATCCAC	824
QY	781	CAGCCCAAGTGCCAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTGAGCCAGGGTCAG	840
DB	825	CAGCCCAAGTGCCAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTGAGCCAGGGTCAG	884
	841	TGCATCTTCAAGAGAGTGGTCAACAACCTCATCCCCTATCTAAAGGACCTCGGCTTGCAT	900
	885	TGCATCTTCAAGAGAGTGGTCAACAACCTCATCCCCTATCTAAAGGACCTCGGCTTGCAT	944
QY	901	TGCCCAACCTTACCAACCCGGCTGACTTCATCATCGAGGTGGCTCTGCGCAGTATGGA	960
DB	945	TGCCCAACCTTACCAACCCGGCTGACTTCATCATCGAGGTGGCTCTGCGCAGTATGGA	1004
	961	GACCTGAACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTGTGCGCTATGGCTGAGAAG	1020
	1005	GACCTGAACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTGTGCGCTATGGCTGAGAAG	1064
QY	1021	AAGAGCAGCCCTGAGAGAAGAGGTCCCTGCCCATGCCCTCCTTGTCTCTCGGAAGTG	1080
DB	1065	AAGAGCAGCCCTGAGAGAAGAGGTCCCTGCCCATGCCCTCCTTGTCTCTCGGAAGTG	1124
	1081	GATCCCATTGAAAGCCACACTTGTGCCACAGACACCTTCACACAGTTTCTGCATCTCTTC	1140
	1125	GATCCCATTGAAAGCCACACTTGTGCCACAGACACCTTCACACAGTTTCTGCATCTCTTC	1184
QY	1141	AAGAGCAGCCCTGTCGTCCATCTCAGGGACAGGTCTGTGACCCACCTTACGGTTTCATGTCC	1200
DB	1185	AAGAGCAGCCCTGTCGTCCATCTCAGGGACAGGTCTGTGACCCACCTTACGGTTTCATGTCC	1244
	1201	CACGTGGTTATTGGCGTGTCTATCGGGCTCCTCTACCTGCTATTTGGCAGCATGCCAGC	1260
	1245	CACGTGGTTATTGGCGTGTCTATCGGGCTCCTCTACCTGCTATTTGGCAGCATGCCAGC	1304
QY	1261	AAGTCTTCAACAACACCGGTGCTCTTCTTCTCCATGCTGTTTCTCTCATGTTTCGGGCC	1320
DB	1305	AAGTCTTCAACAACACCGGTGCTCTTCTTCTCCATGCTGTTTCTCTCATGTTTCGGGCC	1364
	1321	CTCATGCCAATGTGCTCACCTTCCCTTATAGAGATGGCGTCTTTCATGAGGAGCACTTC	1380
	1365	CTCATGCCAATGTGCTCACCTTCCCTTATAGAGATGGCGTCTTTCATGAGGAGCACTTC	1424
QY	1381	AACCTACTGGTACAGCCCTCAAGCGTATTACCTGGCCAAAGACCATGGCTGACGTGCCCTTT	1440
DB	1425	AACCTACTGGTACAGCCCTCAAGCGTATTACCTGGCCAAAGACCATGGCTGACGTGCCCTTT	1484
	1441	CAGTGTGTGTCCGGTGGTCTACTGTGAGCATTTGTACTGAGTACGGGCCAGCCCGCT	1500
	1485	CAGTGTGTGTCCGGTGGTCTACTGTGAGCATTTGTACTGAGTACGGGCCAGCCCGCT	1544
QY	1501	GAGACCAGCCGCTTCTCTGCTCTTCTCAGCCCTGGCCACCGCCCTTGTGTGGCCAA	1560

[illegible]


```
Db 1725 CCCACTTACCTCAATGGAGCTCCTATCTCCTATGTCAGGTATGGCTTTGAGGGTGTG 1784
QY 1741 ATCTTGACCATATGTCGATGAGGAGGAGGAGACCTGACATGTTTAGAGGAACGCTGCCCG 1800
Db 1785 ATCTTGACCATATGTCGATGAGGAGGAGGAGACCTGACATGTTTAGAGGAACGCTGCCCG 1844
QY 1801 TTCCGGAGCCACAGAGCATCTCCGAGGCGCTGGATGTGGAGGATGCCAAGCTCTACATG 1860
Db 1845 TTCCGGAGCCACAGAGCATCTCCGAGGCGCTGGATGTGGAGGATGCCAAGCTCTACATG 1904
QY 1861 GACTTCTCTGTTGGGCATCTTCTTCTAGCCCTGCGGCTGCTGGCCCTACCTTTGTCGTG 1920
Db 1905 GACTTCTCTGTTGGGCATCTTCTTCTAGCCCTGCGGCTGCTGGCCCTACCTTTGTCGTG 1964
QY 1921 CGTTACCCGGGTCAAGTCAAGAGATAGAGCTTGGCCAGCTGTACCCAGCTGTACCCAGCCCTGCA 1980
Db 1965 CGTTACCCGGGTCAAGTCAAGAGATAGAGCTTGGCCAGCTGTACCCAGCTGTACCCAGCCCTGCA 2024
QY 1981 GCAGGAAGCCCCAGTCCAGGCCCTTTGGGACTGTTTAACTTATAGACTTTGGGCACTG 2040
Db 2025 GCAGGAAGCCCCAGTCCAGGCCCTTTGGGACTGTTTAACTTATAGACTTTGGGCACTG 2084
QY 2041 GTTCTTGGGGGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2100
Db 2085 GTTCTTGGGGGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2144
QY 2101 GTTCTTGGGGGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160
Db 2145 GTTCTTGGGGGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2204
QY 2161 TCCCAAGTTGATCGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
Db 2205 TCCCAAGTTGATCGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2264
QY 2221 CTACTGGGAGGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2280
Db 2265 CTACTGGGAGGTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2324
QY 2281 AGCCCTTAGGCTCTTAGGGCCCACTTACAACCTGACCAAGTGGCCCTCTGGGGGTCC 2340
Db 2325 AGCCCTTAGGCTCTTAGGGCCCACTTACAACCTGACCAAGTGGCCCTCTGGGGGTCC 2384
QY 2341 CCACCACACAAGTTTGTAACTGGGCTGCTATAAGTTGGAGTTCCAGGCTGGGCCC 2400
Db 2385 CCACCACACAAGTTTGTAACTGGGCTGCTATAAGTTGGAGTTCCAGGCTGGGCCC 2444
QY 2401 TGGTGGAGTCCACTGGAAGTCCCATATGATGATTTGAAATGGACAGGAGGACTCTGGA 2460
Db 2445 TGGTGGAGTCCACTGGAAGTCCCATATGATGATTTGAAATGGACAGGAGGACTCTGGA 2504
QY 2461 AGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520
Db 2505 AGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2564
QY 2521 GCCAGACAGAAGCT--GGGTTTCTGCTAGGTCACCACTCCCAATCCCT--GGGGATTGGA 2578
Db 2565 GCCAGACAGAAGCTGGGTTTCTGCTAGGTCACCACTCCCAATCCCTGGGGGRTTGA 2624
QY 2579 GAGGCTGGGCTGGGATGCCCATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2638
Db 2625 GRGGCTGGGGTGGGRTGSCCATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2684
QY 2639 CTG 2641
Db 2685 CTG 2687
```

RESULT 6
US-10-090-455-3
; Sequence 3, Application us/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:

```
; APPLICANT: Chen, Hongyun  
; APPLICANT: Le Bihan, Stephane  
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF  
; FILE REFERENCE: 100103.406  
; CURRENT APPLICATION NUMBER: US/10/090,455  
; CURRENT FILING DATE: 2002-03-01  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1941  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-090-455-3  
  
Query Match 56.2%; Score 1941; DB 9; Length 1941;  
Best Local Similarity 100.0%; Pred No. 0;  
Matches 1941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 7 ATGGGGAGAGGCGCTGGAGCCGCTGGGCTCTGGAAGGCGGGGGCTGTGGCCATG 66  
Db 1 ATGGGGAGAGGCGCTGGAGCCGCTGGGCTGTGGACTAGGGCGGGGGCTGTGGCCATG 60  
QY 67 GCGGTGACCTGGAGACGCGGGGCGGAACCCCTGTGCTGACACGACACCTGAAGAAGTG 126  
Db 61 GCGGTGACCTGGAGACGCGGGGCGGAACCCCTGTGCTGACACGACACCTGAAGAAGTG 120  
QY 127 GAGAACCATCACTCAAGCCCAAGCGCTTCTCCACCTGCCAAGCGCTCAGCCGTGGAC 186  
Db 121 GAGAACCATCACTCAAGCCCAAGCGCTTCTCCACCTGCCAAGCGCTCAGCCGTGGAC 180  
QY 187 ATCGAGTTCTGGAGCTGTCTATTCTCGTGGGGAGGGGCCCTGTGGCGCAAAAGGGGT 246  
Db 181 ATCGAGTTCTGGAGCTGTCTATTCTCGTGGGGAGGGGCCCTGTGGCGCAAAAGGGGT 240  
QY 247 TATAAGACCTTCTCAAGTGCCTCTCAGGTAAATTTCTGGCGGGGAGCTGATGGCATC 306  
Db 241 TATAAGACCTTCTCAAGTGCCTCTCAGGTAAATTTCTGGCGGGGAGCTGATGGCATC 300  
QY 307 ATGGGGCCCTCAGGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 366  
Db 301 ATGGGGCCCTCAGGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 360  
QY 367 TCTGGAATGAAGGGGAGATCTCTGTTAATGGAAGGCCACGGGAGCTGAGGACCTTCCCG 426  
Db 361 TCTGGAATGAAGGGGAGATCTCTGTTAATGGAAGGCCACGGGAGCTGAGGACCTTCCCG 420  
QY 427 AGATGCTCTCTACATCATGCAAGTGCATCTGCTGCCGACACCTCAGGTTGGAA 486  
Db 421 AGATGCTCTCTACATCATGCAAGTGCATCTGCTGCCGACACCTCAGGTTGGAA 480  
QY 487 GCATGATGCTCTCTCTAACCTGAATCTTACTGAGAATCCCGATGTGAAAAACGATCTC 546  
Db 481 GCATGATGCTCTCTCTAACCTGAATCTTACTGAGAATCCCGATGTGAAAAACGATCTC 540  
QY 547 GTGACAGATCTGACGGCACTGGCCCTGATGCTGCTGCCACAGGAGGAGCCCTG 606  
Db 541 GTGACAGATCTGACGGCACTGGCCCTGATGCTGCTGCCACAGGAGGAGCCCTG 600  
QY 607 CTCTCTGGCGGAGAGGAGCTGTGGCCATCGCCCTGGAGCTGTCAACACCCGCT 666  
Db 601 CTCTCTGGCGGAGAGGAGCTGTGGCCATCGCCCTGGAGCTGTCAACACCCGCT 660  
QY 667 GTCATGCTCTTCTGATGAGCCACAGTGGTCTGGATAGCCCTTGTGTTCCAAAGTGGT 726  
Db 661 GTCATGCTCTTCTGATGAGCCACAGTGGTCTGGATAGCCCTTGTGTTCCAAAGTGGT 720  
QY 727 TCCCTCATGAAGTCCCTGGCACAGGGGGCCGTACCATCATCTGSCACCATCCACAGCCC 786  
Db 721 TCCCTCATGAAGTCCCTGGCACAGGGGGCCGTACCATCATCTGSCACCATCCACAGCCC 780  
QY 787 AGTGCCAAAGCTCTTTGAGATGTTTGAAGCTCTACATCTCTGAGCCAGGCTCAGTGCATC 846  
Db 781 AGTGCCAAAGCTCTTTGAGATGTTTGAAGCTCTACATCTCTGAGCCAGGCTCAGTGCATC 840
```



```
QY 847 TTCAAAGAGTGTGTACCAACCTGATCCCTATCTAAAGGACTCGCTTGCATTGCCCC 906
Db 841 TTCAAAGAGTGTGTACCAACCTGATCCCTATCTAAAGGACTCGCTTGCATTGCCCC 900
QY 907 ACCTACACAAACCCGGCTGACTTCATCATCGAGGTGGCTCTCGCGAGTATGAGACCTG 966
Db 901 ACCTACACAAACCCGGCTGACTTCATCATCGAGGTGGCTCTCGCGAGTATGAGACCTG 960
QY 967 AACCCCATGTTGTTCAAGGCTGTCAGAAATGGCTGTGGCTATGCTGAGAGAGAGC 1026
Db 961 AACCCCATGTTGTTCAAGGCTGTCAGAAATGGCTGTGGCTATGCTGAGAGAGAGC 1020
QY 1027 AGCCCTGAGAAGAACAGGTCCCTGCCCCATGCCCTCTTGTCTCCGGAAGTGGATCCC 1086
Db 1021 AGCCCTGAGAAGAACAGGTCCCTGCCCCATGCCCTCTTGTCTCCGGAAGTGGATCCC 1080
QY 1087 ATTGAAGGCACACCTTTTGGCAGCAGACCCCTCACACAGTTTCATCTCTTCAAGAGG 1146
Db 1081 ATTGAAGGCACACCTTTTGGCAGCAGACCCCTCACACAGTTTCATCTCTTCAAGAGG 1140
QY 1147 ACCTTCTCTGTCATCTCAGGACAGCGTCCCTGACCCAGCTACGGTTCATGTCCACGCTG 1206
Db 1141 ACCTTCTCTGTCATCTCAGGACAGCGTCCCTGACCCAGCTACGGTTCATGTCCACGCTG 1200
QY 1207 GTTATTGGGCTGCTCATCGGCTCCCTCTACCTGCATATTGGCAGCATGCCAAGGTC 1266
Db 1201 GTTATTGGGCTGCTCATCGGCTCCCTCTACCTGCATATTGGCAGCATGCCAAGGTC 1260
QY 1267 TTCAAACACACCGGCTGCTCTTCTTCATGCTGTTCTCATGTTTCGCGCCCTCATG 1326
Db 1261 TTCAAACACACCGGCTGCTCTTCTTCATGCTGTTCTCATGTTTCGCGCCCTCATG 1320
QY 1327 CCAACTGTGCTACCTTCCTCTTAGAGATGGGCTCTTCATGAGGAGCAGCTCAACTAC 1386
Db 1321 CCAACTGTGCTACCTTCCTCTTAGAGATGGGCTCTTCATGAGGAGCAGCTCAACTAC 1380
QY 1387 TGTACAGGCTCAAAAGCGTATTACCTGGCCAGACCATGGCTGACCTGGCCCTTTCAGGTG 1446
Db 1381 TGTACAGGCTCAAAAGCGTATTACCTGGCCAGACCATGGCTGACCTGGCCCTTTCAGGTG 1440
QY 1447 GTGTGTCGGTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCTGAGACC 1506
Db 1441 GTGTGTCGGTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCTGAGACC 1500
QY 1507 AGCGCTTCTGCTCTCTCTCAGCCCTGGCCACCGCCAGCCCTTGTGGCCCAATCTTG 1566
Db 1501 AGCGCTTCTGCTCTCTCTCAGCCCTGGCCACCGCCAGCCCTTGTGGCCCAATCTTG 1560
QY 1567 GGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGTGGCCACTTTTGTGGGCCAGTT 1626
Db 1561 GGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGTGGCCACTTTTGTGGGCCAGTT 1620
QY 1627 AGCGCATCCCTGCTCTCTTCTCTCGGCTTCTTTGTGAGCTTCAAGACCATTCGCCACT 1686
Db 1621 AGCGCATCCCTGCTCTCTTCTCTCGGCTTCTTTGTGAGCTTCAAGACCATTCGCCACT 1680
QY 1687 TACTGCAATGAGCTTCTCTCTCTATGTGAGTATGGCTTTGAGGCTGTGATCTCG 1746
Db 1681 TACTGCAATGAGCTTCTCTCTCTATGTGAGTATGGCTTTGAGGCTGTGATCTCG 1740
QY 1747 AGCATCTATGGATGGAGGAGACCTGACATGTTTATAGGAAGCGTCCCGTTCGG 1806
Db 1741 AGCATCTATGGATGGAGGAGACCTGACATGTTTATAGGAAGCGTCCCGTTCGG 1800
QY 1807 GAGCCACAGCATCTCCGAGCGTGGATGTGGAGATGCCAAGCTCTACATGACTTC 1866
Db 1801 GAGCCACAGCATCTCCGAGCGTGGATGTGGAGATGCCAAGCTCTACATGACTTC 1860
QY 1867 CTGGCTTGGGATCTTCTCTAGCCCTGCGGCTGCTGCGCTACCTTGTGCTGGTTAC 1926
Db 1861 CTGGCTTGGGATCTTCTCTAGCCCTGCGGCTGCTGCGCTACCTTGTGCTGGTTAC 1920
```

```
QY 1927 CGGTCAAGTCAGAGAGATAG 1947
Db 1921 CGGTCAAGTCAGAGAGATAG 1941

RESULT 7
US-10-037-270-918
; Sequence 918, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungling
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US/10/037,270
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 918
; LENGTH: 3376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2808)
US-10-037-270-918

Query Match 30.7%; Score 1062.4; DB 9; Length 3376;
Best Local Similarity 81.5%; Pred. No. 1e-298;
Matches 1389; Conservative 0; Mismatches 21; Indels 294; Gaps 2;

QY 227 CCTGCTGGCGCAAAAGGGGTTATAGACCCCTTCTCAAGTGCCTCTCAGGTAATTTCTGCC 286
Db 38 CGCAGGTCACACAGCTGTTTATAGACCCCTTCTCAAGTGCCTCTCAGGTAATTTCTGCC 97
QY 287 GCGCGGAGCTGATTTGGCATCATCGGGCTCGGAGGCTGCAAGTCTACATTCATCAACA 346
Db 98 GCGGGAGCTGATTTGGCATCATCGGGCTCGGAGGCTGCAAGTCTACATTCATCAACA 157
QY 347 TCTTGGCAGGATACAGGGAGTCTGGAATGAAGGGGAGATCTCTGTTAATGAAGGCCAC 406
Db 158 TCTTGGCAGGATACAGGGAGTCTGGAATGAAGGGGAGATCTCTGTTAATGAAGGCCAC 217
QY 407 GGGAGCTGAGGACCTTCCCGAAGATGTCCTGCTACATCATGCAAGATGACATGCTGCTGC 466
Db 218 GGGAGCTGAGGACCTTCCCGAAGATGTCCTGCTACATCATGCAAGATGACATGCTGCTGC 277
QY 467 CGCACCTACCGTGTGTGGAGGATGATGCTCTCTGCTAACCTGATCTTACTTGAGATC 526
Db 278 CGCACCTACCGTGTGTGGAGGATGATGCTCTCTGCTAACCTGATCTTACTTGAGATC 337
QY 527 CCGATGTGAAAAACGATCTGTCGACAGAGATCTGACAGGACCTGGGCGCTGATGCTGCT 586
Db 527 CCGATGTGAAAAACGATCTGTCGACAGAGATCTGTCGACAGGACCTGGGCGCTGATGCTGCT 586
```

Db 338 AGGAGGTGAAGAGGAGCTGGTGACAGAGATCCTGACGGCAGCTGGCCCTGATGTCGTGCT 397
Qy 587 CCCACAGGAGCAGCCCTGCTCTCTGGGGGAGAGAGCGCTCTGGCCATCGCCCTGG 646
Db 398 CCCACAGGAGCAGCCCTGCTCTCTGGGGGAGAGAGCGCTCTGGCCATCGCCCTGG 457
Qy 647 AGCTGGTCAACACCCGCTGCTGATGCTCTTTGATGAGCCACAGTGGTCTGGATAGCG 706
Db 458 AGCTGGTCAACACCCGCTGCTGATGCTCTTTGATGAGCCACAGTGGTCTGGATAGCG 517
Qy 707 CTTCTTTGTTTCCAAAGTGGTCTCTCATGAACTCCCTGGCAGAGGGGGCCGTACCATCA 766
Db 518 CTTCTTTGTTTCCAAAGTGGTCTCTCATGAACTCCCTGGCAGAGGGGGCCGTACCATCA 577
Qy 767 TCTGCACATCCACAGCCAGTGCACAGCTCTTTGAGATGTTTGACAAGCTCTACATCC 826
Db 578 TCTGCACATCCACAGCCAGTGCACAGCTCTTTGAGATGTTTGACAAGCTCTTTGACA 625
827 TGAGCCAGGCTAGTGCATCTTCAAGAGAGTGTCAACCACTGATCCCTATCTAAAGG 886
626 -----AGTGCATCTTCAAGAGCGTGGTCAACCACTGATCCCTATCTAAAGG 673
Qy 887 GACTCGGCTGCAATGCCCCACCTACACAAACCCGGTGAATTCATCATCGAGGTGGCT 946
Db 674 GACTCGGCTGCAATGCCCCACCTACACAAACCCGGTGAATTCATCATCGAGGTGGCT 733
Qy 947 CTGGCAGTATGAGAGCCTGAACCCCATGTTGTTGAGGCTGTGAGAGTGGCTGTGCG 1006
Db 734 CTGGCAGTATGAGAGCCTGAACCCCATGTTGTTGAGGCTGTGAGAGTGGCTGTGCG 793
Qy 1007 CTATGGCTGAGAAGAGAGCAGCCCTGAGAAGAACAGGTCCTGCCCCATGCCCTCCCT 1066
Db 794 CTATGGCTGAGAAGAGAGCAGCCCTGAGAAGAACAGGTCCTGCCCCATGCCCTCCCT 853
Qy 1067 GTCTCCGGAGTGGATCCCATTTGAAGCCACACCTTTGCCACAGCAGCCCTCACAGT 1126
Db 854 GTCTCCGGAGTGGATCCCATTTGAAGCCACACCTTTGCCACAGCAGCCCTCACAGT 913
Qy 1127 TCTGCATCTCTTCAAGAGGAGCTTCTGTCATCTCTCAGGAGCAGGCTCTGACCCACC 1186
Db 914 TCTGCATCTCTTCAAGAGGAGCTTCTGTCATCTCTCAGGAGCAGGCTCTGACCCACC 959
Qy 1187 TACGGTTTCATGTCACAGGCTTATTGAGGCTGCTCATCTGAGGCTCTCTCTACCTGATATG 1246
Db 960 ----- 959
1247 GCGAGCATGCCAAGGCTTCAACACACACCGGCTGCTCTCTCTCTCTCTCTCTCTCT 1306
Db 960 ----- 959
Qy 1307 TCATGTTGCGCGCCCTCATGCCAAGTGTGCTCACCTTCCCTTAGAGATGGCGGTCTCA 1366
Db 960 ----- 959
Qy 1367 TGAGGAGACCTCAACTACTGTGTACAGCCTCAAGCGTATTACCTGGCCAAAGCATTGG 1426
Db 960 ----- 959
Qy 1427 CTGACGTGCCCTTTCAGGTGGTGTCCGGTGGTCTACTCAGCATTTGTACTGGATGA 1486
Db 960 -----GGTGGTGTCCGGTGGTCTACTCAGCATTTGTACTGGATGA 1003
Qy 1487 CGGCCAGCCGCTGAGACAGCGCTTCTGCTCTCTCTCAGGCTCTGCGCCACGCGCCACCG 1546
Db 1004 CGGCCAGCCGCTGAGACAGCGCTTCTGCTCTCTCTCAGGCTCTGCGCCACGCGCCACCG 1063
Qy 1547 CCTTGTGGCCCAATCTTTGGGCTGCTGATCGAGAGTGTCTTCAACTCCCTACAGGTGG 1606
Db 1064 CCTTGTGGCCCAATCTTTGGGCTGCTGATCGAGAGTGTCTTCAACTCCCTACAGGTGG 1123
Qy 1607 CCACCTTTGTGGCCCAATCTTACGCCATCCCTGCTCTCTTGTCTCGGCTCTTCTTCTCA 1666
Db 1124 CCACCTTTGTGGCCCAATCTTACGCCATCCCTGCTCTCTTGTCTCGGCTCTTCTTCTCA 1183

RESULT 8

US-09-954-531-591
; Sequence 591, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 591
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-591

Query Match 28.28; Score 972.8; DB 9; Length 2930;
Best Local Similarity 70.5%; Pred. No. 1.2e-272;
Matches 1351; Conservative 0; Mismatches 517; Indels 48; Gaps 2;

Qy 81 GGACGGGGCGAAGCCCTGCTGACGACGACCTGAAGAGGTGGAGAACACATCAC 140
Db 141 GGAGCCACTGAGACGACCTGCTGANTGACATCTGAAAAGTAGATATACCTCAC 200
Qy 141 TGAAGCCAGCGTCTTCCACCTGCGCAAGCGCTCAGCCGTGGACATCGAGTTCTGGA 200
Db 201 GGAAGCCAGCGTCTTCCCTCTGCTCGGAGGCGAGCTCTGAACATTGAATTCAGGA 260
Qy 201 GCTGCTCTATTCGCTGGGGAGGGCCCTGCTGGCGCAAGAGGGTTATAGACCTTCT 260
Db 261 CTTTCTCTATTCGCTGCTGAGGACCTGCTGAGGAGGAGGAGGAGGAGGAGGAGG 320
Qy 261 CAAGTGCCTCTCAGGTAAATTCCTGCCCGGAGCTGATTTGGCATCATGGCCCTCAGG 320
Db 321 GAAAGGAATTCGCGGAGGTTCAATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 380
Qy 321 GGCTGGCAAGTCTAGATTCATGAACATCTTTGGCAGGATACAGGAGGCTCTGGAATGAAGG 380

Db 381 GCGCGGGAAGTCCACGCTGATGAACATCTCTGGCTGATACAGGAGACGGGCAATGAAGG 440
QY 381 GCAGATCCTGTTAATGAAGAGCCACGGAGCTGAGGACCTCCCGAAGATGCTCTGCTA 440
Db 441 GCGCGTCTCATCAAGCGGCTGCCCGGACCTGCGCTCTCCGGAAGGTGCTCTGCTA 500
QY 441 CATCATGCAAGATGACATGCTGCTGCGGACCTCAGCGGTGTTGGAAGCCATGATGCTCTC 500
Db 501 CATCATGCAAGATGACATGCTGCTGCGGACCTCAGCTGTCAGAGGCGCATGATGCTGTC 560
QY 501 TGTACCTGATCTTACTGAGAAATCCCGATGTAAGAGATCTCGTGACAGAGATCCT 560
Db 561 GGCACATCTGAAGCTTCAGAGAGAGATGAAGCAGAGAGAAATGCTCAAGGATACT 620
QY 561 GAGCGACCTGGGCTGATGCTGCTGCGGACCTCAGCGGTGTTGGAAGCCATGATGCTCTC 620
Db 621 GACAGCGTGGGCTGCTGCTTTCGCGCAACAGCGGACCGGAGCGCTGTCAGGTGGTCA 680
QY 621 GAGGAAGCGTCTGGCCATCCCTGGAGCTGCTCAACAAACCCGCTGTCATGCTCTTGA 680
Db 681 GCGCAAGCGCTGGCCATCGCGTGGAGCTGGTGAACAACCCCTCCAGTCATGCTTTCGA 740
QY 681 TGAGCCACAGTGGTGGATAGAGCGCTCTGTTTCCAAAGTGGTGTCTCCATGATGATC 740
Db 741 TGAGCCACAGCGGCTGGACAGCGCTCTGTTTCCAAAGTGGTGTCTCCATGATGATC 800
QY 741 CTGGCACAGCGGCGGCTGACATCATCTGCACATCCACAGCGGCTGCTCAAGCTCTT 800
Db 801 GCTCGCTCAAGGGGTGCTCCATCATTTGCAACCTCCACAGCGGCTGCTCAAGCTCTT 860
QY 801 TGAGATGTTGCAAGCTCTACATCTGAGCCAGGCTCAGTGCATCTTCAAGAGTGGT 860
Db 861 CGAGCTGTTGCAAGCTCTACATCTGAGTCAAGGACATGCTGCTACCGGGGAAAGT 920
QY 861 CACCAACCTGATCCCTATCTAAGGAGCTGCGCTTGCATTTGCCCCACCTACCAACCC 920
Db 921 CTGCAATCTGTGCCATATTTAGGAGTTTGGGATTTGGAACCTGCCAACCTACCAACCC 980
QY 921 GGCTGACTCATCATGAGTGGGCTCTGCGAGTATGAGACCTCAACCCCATGTTGT 980
Db 981 AGCAGATTTGTCATGAGTGGTTCATCCGCGAGTACGGTGATCAGAACAGTGGTGGT 1040
QY 981 CAGGCTGTGCAAGATGAGGCTGTCGCTATGCTGAGAGAGAGAGAGCCCTGAGAGAA 1040
Db 1041 GAGAGCGGTTCGGAGGAGTGTGACTCAGACCAACAGAGAGACCTCGGGGTGATGC 1100
QY 1041 CGAGGTCCTGCCCCATGCCCTTGTCTCCGGAAGTGGATCCCATTCARA----- 1093
Db 1101 CGAGGTGAACCTTTTCTTTGGCAGCGCCCTCTGAAGAGGTAAAGCAGACAAACGATT 1160
QY 1094 -----GCCACACCTTTGCCACCAACGAC 1115
Db 1161 AAAGGGTTGAGAAAGGACTCTCTGTCATGGAAGCTGCCACAGCTTCTCTGCCAGCTG 1220
QY 1116 CTTACACAGTCTGATCCTCTTCAAGAGGACCTTCTGTCATCCTCAGGAGACAGGT 1175
Db 1221 CTTACAGGATTTGATCCTCTTCAAGAGGACCTTCTCAGCATCATGAGGAGCTCGGT 1280
QY 1176 CTTGACCCACCTTACGTTGATGTCACAGCTGTTATTGGCTGCTCATCGGCTCCTCTA 1235
Db 1281 CTTGACACACCTGCGCATCCTCGCACATTTGGATCGGCTCTCTCTGCTGTTA 1340
QY 1236 CTTGATATTGGCAGATGCGCAGGATCTTCAACACACCGGCTGCTCTCTCTCTC 1295
Db 1341 CTTGGGATCGGAAGCAACCAAGAGTCTTGAAGACTCGGCTCTCTCTCTCTCTC 1400
QY 1296 CATGCTGTTCTCATGTTGCGCGCTCTCATGCGCACTGCTGCTACCTTCCCTTTAGAT 1355
Db 1401 CATGCTGTTCTCATGTTGCGCGCTCTCATGCGCACTGCTGCTGCTGCTGCTGCTG 1460
QY 1356 GCGGCTCTCATGAGGAGCCTCAACTACTGTTGATGCTGCTGCTGCTGCTGCTGCTG 1415
Db 1461 GCGAGTCTTCTTGGGAAACCTCAACTACTGTTGATGCTGCTGCTGCTGCTGCTGCTG 1520

QY 1416 CAAGACCATGGCTGACCTGCCCTTTTCAGGTGGTGTGTCGGTGTCTACTGACGATGTT 1475
Db 1521 CAAGACCATGGGAGAGCTGCCCTTTTCAGATCATGTTCCAGTGGCTACTGACGATCGT 1580
QY 1476 GTACTGGATGACGGGCGGAGCCGCTGAGACACCGCTTCTCTCTCTCTCTCTCTCT 1535
Db 1581 GTACTGGATGACGCTCGACGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1640
QY 1536 CACCGCCACCGCTTGGTGGCCCAATCTTTGGGGCTGCTGATGCGGAGCTGCTTCCAACTC 1595
Db 1641 CACCATGACCTCTCCCTGGTGGCACAGTCCCTGGGCTGCTGATGCGAGCGCTCCACGTC 1700
QY 1596 CTTTACAGTGGGCGGCTTTTGGGCGGAGTACCGGATCCCTGCTCTCTCTCTCTCTCT 1655
Db 1701 CTTGAGTGGGCGGCTTTTGGGCGGAGTACCGGATCCCTGCTCTCTCTCTCTCTCT 1760
QY 1656 CTTTCTTTGTCAGCTTCAAGACCATCCCACTTACCTGCAATGAGGCTCTCTCTCTCT 1715
Db 1761 GTTCTCGTCACTTCGACACCATCCCACTACCTACAGTGGATGCTCTACATCTCTCTA 1820
QY 1716 TGTACAGTATGGCTTTGAGGGTGTGATCTGACGATCTATGAGTGGAGGAGGACCT 1775
Db 1821 TGTACAGTATGGCTTTGAGGGTGTGATCTGACGATCTATGAGTGGAGGAGGACCT 1880
QY 1776 GACATGT---TTAGAGGAACGCTTGGCGGAGCCACAGAGCATCTCTCGAGCGCT 1832
Db 1881 GCACTGTGACATCGACGACGCTGCCACTTCCAGAGTGGAGGAGGAGGAGCT 1940
QY 1833 GGATGTTGAGGATGCGCAAGCTCTACATGAGCTTCTCTGCTTGGGATCTCTCTCTCT 1892
Db 1941 GGAGTGGAAATGCGCAAGCTGTACCTGAGCTTCTCTGAGTCTGCGGATTTCTCTCTCT 2000
QY 1893 CTTGCGGCTGCTGGCTTACCTTGTGCTGCTTACCGGGTCAAGTCAAGAGAGATAGA 1948
Db 2001 CTTGCGGCTTGTGCTTATTTGGTCTCTAGGTACAAAAATCCGGGAGAGAGGTAA 2056

RESULT 9

US-10-171-581-276
; Sequence 276, Application US/10171581
; Publication No. US2003010426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10171581
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 276
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLIC INFORMATION:
; DATABASE ACCESSION NUMBER: X91249
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-276

Query Match 28.2%; Score 972.8; DB 9; Length 2930;

Best Local Similarity 70.5%; Pred No. 1.2e-272;

Matches 1351; Conservative 0; Mismatches 517; Indels 48; Gaps 2;

QY 81 GGAGCGGCGGAGACCCCTTGTGCTGACACGACCTTCAAGAGGTGGAGAACCATCATC 140
Db 141 GGAGCGCACTGAGAGGACCTGCTGAATGAGATCTGAAAAAGTAGATAAATACCTCAC 200
QY 141 TGAAGCCGAGCGCTTCTCCACCTGCCAAGCGCTGACCGCTGGAGATCGAGTTGTTGA 200
Db 201 GGAAGCCGAGCGCTTCTCTCTCTGCTCGGAGGCGCTGTGAACATTGAATTCAGGGA 260

Db 431 GGAACCCCTGTGCTGACCAACGACCTGAAGAGGTGGAGAACCAACATCACTGAAGCCCA 372
QY 150 GCGCTTCTCCCACTCCCAAGGCTCAGCCGTGGACATCGAGTTTCGTGGAGCTGCTCA 209
Db 371 GCGCTTCTCCCACTCCCAAGGCTCAGCCGTGGACATCGAGTTTCGTGGAGCTGCTCA 312
QY 210 TTCCTGCGGGAGGGGCGCTGCTGGCGCAAAAGGGGT 246
Db 311 TTCCTGCGGGAGGGGCGCTGCTGGCGCAAAAGGGGT 275

RESULT 13
US-09-989-981A-5
; Sequence 5, Application US/0998981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-5

Query Match 3.1%; Score 107.2; DB 9; Length 2340;
Best Local Similarity 49.5%; Pred. No. 1.2e-20;
Matches 379; Conservative 0; Mismatches 363; Indels 24; Gaps 3;

QY 212 CCGTGGGAGGGGCGCTGCTGGCGCAAAAGGGGTATAGACCCTTCTCAAGTCCCTCT 271
Db 264 CCGTGGGAGCATCATCTTCCCGGACAGCTGGACCGAGATCTCTCAAGATGTCT 323
Db 272 CAGGTAATTTCTGCCCGGGAGCTGATGGCATCATGGCCCTCAGGGCTGGCAACT 331
Db 324 CTTGTACGTGGAGCGGGAGATCATGTGCTAGGAGCTCAGGCTCCGGGAAA 383
QY 332 CTACATTCATGAACATCTTGGCAGGATACAGGAGTCTGGAATGA-----AGGGCAGA 385
Db 384 CCACGCTGTGGAGCCATGTCGGGAGGCTGGGCGCGCGGGACCTTCTCTGGGGAGG 443
QY 386 TCTGTTAATGAAGGCCAGGGAGCTGAGGACCTTCGCGAAGATGCTCCTGATACGA 445
Db 444 TGTATGTGAACGCGGGCGCTCGCGGAGAGCTTCCAGGACGTCTCTCTAGCTCC 503
QY 446 TCAAGATCACATGCTGCTGGCGACCTCACCGTGTGGAGCCATGATGGTCTCTGCTA 505
Db 504 TCGAGGCGACACCTCTGAGCAGCTCACCCTCGCGGCGAGAGCTGCACTACCGCGC 563
QY 506 ACCTG---AATCTTACTGAGATCTCCGATGTGAAACACATCTCGTGACAGAGATCCTGA 562
Db 564 TCTGCGCATCCCGCGCAATCCGGCTCTTCCAGAGAGGTGGAGCGGCTCATGG 623
QY 563 CGGACTGGGCTGATGCTGCTCCACACAGACAGACCCCT-----GC 607
Db 624 CAGAGCTGAGTGTAGCCATGTGGCAGACCGAGCTGATTGGCAACTACAGCTTGGGGGCA 683

QY 608 TCTCTGGCGGAGAGAGCGCTGCGCCATCGCCCTGGAGCTGGTCAACAACCCGCTG 667
Db 684 TTTCACAGGCTGAGCGGCGCCGGGTCTCCATCGAGCCAGCGTGTCTCAGGATCTTAAGG 743
QY 668 TCATGTTCTTTGATGAGCCACAGTGGTCTGATAGCGCTCTTTGTTTCCAAAGTGGTCT 727
Db 744 TCATGCTGTTGATGAGCCACACAGCGCTGGATGCTGCTATATCAGATTGTCTG 803
QY 728 CCCTCATGAAGTCCCTGGCAGAGGGGCGGTACCATCATCTGCACCATCCACAGCCCA,787
Db 804 TCCTCTCTGTGAACCTGGCTCGCAGGAACGAATGTGTCTCAGCATTCACAGCCCC 863
QY 788 GTGCCAAGCTCTTGGAGATGTTTGACAGCTCTACATCTGAGCAGGCTCAGTGCATCT 847
Db 864 GTTCTGAGCTTTTTCAGCTCTTTTGACAAAATTTGCCATCTCGGAGAGCTGATTT 923
QY 848 TCAAGAGAGTGTTCACCAACCTGATCCCCCTATCTAAAGGAGCTCGGCTTGCATTGCCCA 907
Db 924 TCTGTGGCAGCGCAGCGGAATGCTTGATTTCTTCAATGACTGCGGTTACCCCTTGCTCTG 983
QY 908 CTTACACAAACCGGCTGACTCATCATCATGAGGTGGCTCTTGCGGA 953
Db 984 AACATTCAAACCTTTTGACTTCTATATGAGCTGACCTGACGTCACTGGA 1029

RESULT 14
US-09-837-992-4
; Sequence 4, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human sitosterolemia gene (SSG)
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: protein
US-09-837-992-4

Query Match 3.1%; Score 107.2; DB 10; Length 2340;
Best Local Similarity 49.5%; Pred. No. 1.2e-20;
Matches 379; Conservative 0; Mismatches 363; Indels 24; Gaps 3;

QY 212 CCGTGGGAGGGGCGCTGCTGGCGCAAAAGGGGTATAGACCCTTCTCAAGTCCCTCT 271
Db 264 CCGTGGGAGCATCATCTTCCCGGACAGCTGGACCGAGATCTCTCAAGATGTCT 323
QY 272 CAGGTAATTTCTGCCCGGGAGCTGATGGCATCATGGCCCTCAGGGCTGGCAACT 331
Db 324 CTTGTACGTGGAGCGGGAGATCATGTGCTAGGAGCTCAGGCTCCGGGAAA 383
QY 332 CTACATTCATGAACATCTTGGCAGGATACAGGAGTCTGGAATGA-----AGGGCAGA 385
Db 384 CCACGCTGTGGAGCCATGTCGGGAGGCTGGGCGCGCGGGACCTTCTCTGGGGAGG 443

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:21:33 ; Search time 164 seconds
(without alignments)
6460.782 Million cell updates/sec

Title: US-10-072-621-4
Perfect score: 3455
Sequence: 1 gccaccatggcggaagc.....atgccagttgtattgtggg 3455
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

ched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Resub Patents_NA:
1: /cn2.6/prodata/1/ina/5A_COMB.seq:
2: /cn2.6/prodata/1/ina/5A_COMB.seq:
3: /cn2.6/prodata/1/ina/6A_COMB.seq:
4: /cn2.6/prodata/1/ina/6B_COMB.seq:
5: /cn2.6/prodata/1/ina/PTUS_COMB.seq:
6: /cn2.6/prodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.2	2.3	2418	4	US-09-245-808-2
2	75	2.2	4403765	4	US-09-103-840A-2
3	75	2.2	4411529	4	US-09-103-840A-1
4	73	2.1	7218	1	US-08-232-463-14
5	69	2.0	6909	4	US-09-199-637A-111
6	63.2	1.8	1518	2	US-08-997-080-88
7	63.2	1.8	1518	2	US-08-997-362-88
8	63.2	1.8	1518	3	US-08-873-970-88
9	63.2	1.8	1518	4	US-09-095-855-88
10	63.2	1.8	1518	4	US-09-324-542-88
11	63.2	1.8	1518	4	US-09-205-426-88
12	62.8	1.8	4403765	4	US-09-103-840A-2
13	62.8	1.8	4411529	4	US-09-103-840A-1
14	61.4	1.8	1926	4	US-09-249-585A-4
15	61.4	1.8	1931	2	US-09-130-114-2
16	59.2	1.7	11672	4	US-09-441-340-2
17	57.8	1.7	20986	4	US-08-961-527-54
18	57.4	1.7	16593	4	US-08-961-527-52
19	57.2	1.7	15611	4	US-09-441-340-1
20	56.8	1.6	13121	4	US-08-961-527-126
21	56.6	1.6	5045	4	US-09-390-721-1
22	56.6	1.6	5045	4	US-09-390-721-3
23	54.6	1.6	13440	4	US-08-961-527-128
24	51.8	1.5	1460	4	US-08-858-207A-49
25	51.8	1.5	8145	4	US-08-961-527-122
26	51.6	1.5	23673	4	US-08-773-816-1
27	50.6	1.5	3489	2	US-08-728-323A-1

c	28	50.6	1.5	3489	4	US-09-298-568-1	Sequence 1, Appli
	29	50.6	1.5	32207	2	US-08-770-379-20	Sequence 20, Appl
	30	50.6	1.5	32207	4	US-08-757-669A-20	Sequence 20, Appl
	31	50.6	1.5	32207	4	US-09-230-371A-20	Sequence 20, Appl
	32	49.2	1.4	732	4	US-08-919-573-3	Sequence 3, Appli
	33	49.2	1.4	735	4	US-08-919-573-1	Sequence 1, Appli
	34	48.8	1.4	28804	2	US-08-592-874-1	Sequence 1, Appli
	35	48.8	1.4	28804	3	US-09-096-942-2	Sequence 2, Appli
	36	48.8	1.4	28804	3	US-09-096-867-2	Sequence 4, Appli
	37	48.2	1.4	1959	4	US-09-061-764A-4	Sequence 2, Appli
	38	48.2	1.4	2061	4	US-09-061-764A-17	Sequence 17, Appl
	39	47.8	1.4	966	2	US-08-766-738-2	Sequence 2, Appli
	40	47.8	1.4	966	4	US-09-262-610-2	Sequence 2, Appli
	41	47.8	1.4	4635	4	US-09-221-017B-187	Sequence 187, App
	42	47.2	1.4	6854	4	US-09-194-905-7	Sequence 7, Appli
	43	47	1.4	25002	4	US-08-961-527-48	Sequence 48, Appl
	44	46.2	1.3	5894	3	US-08-665-259-24	Sequence 24, Appl
	45	46.2	1.3	5894	3	US-08-762-500-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-09-245-808-2
; Sequence 2, Application US/09245808
; Patent No. 6313277
; GENERAL INFORMATION:
; APPLICANT: Doyle, L. Austin
; APPLICANT: Abruzzo, Lynne V.
; APPLICANT: Ross, Douglas D.
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which
; FILE OF INVENTION: encodes it
; FILE REFERENCE: Ross Umb conversion
; CURRENT APPLICATION NUMBER: US/09/245,808
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/073763
; EARLIER FILING DATE: 1998-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-2

Query Match	2.3%	Score 79.2;	DB 4;	Length 2418;
Best Local Similarity	47.4%	Pred. NO. 2.4e-10;		
Matches	323;	Conservative	0;	Mismatches 338; Indels 21; Gaps 2;
QY	295	CTGATTGGCATCATGGGCCCTCAGGGGCTGGCAAGTCTACATTCATGAACATCTTGGCA	354	
Db	461	CTCAAGCCATCTCTGGGACCCAGGTTGAGGCAATCTTCGTTATTAGATGCTTAGCT	520	
QY	355	GGATACAGGGAGTCTGGAATGAAGGGGAGATCCTGTTAATGAAGGCCAGGGAGCTG	414	
Db	521	GCAAGGAAAGATCCAAGTGGATTTCTGGAGATGTTCTGATAAATGGAGCAGCGACCT	580	
QY	415	AGGACCTTCGGCAAGATGCTCTGCTACATCATCAAGATGATGCTGCTGCGCCACCTC	474	
Db	581	GCCAAATTTCAAATGTAAATTCAGGTTACGTTGATGATGATGTTGATGGGCACTCTG	640	
QY	475	ACGGTGTGGAGCCCATGATGCTCTGCTGCAACCTGAATCTT-----ACTGAGAAATCCC	528	
Db	641	ACGGTGAAGAAACCTTACAGTTCTCAGAGCTTCTCGGCTTCCACAACTATGACGAAT	700	
QY	529	GATGTGAAACCAATCTCGTGACAGATGCTTCAAGGCACTGCGGCTGATGCTGCTGCC	588	
Db	701	CATGAAACCAACCAAGGATTAACAGGTCATTCAAGAGTTAGTCTGGAATAAGTGCA	760	
QY	589	CACACAGGACACCCCTGC-----TCTCTGGCGGCGAGAGGAGCGCTG	633	
Db	761	GACTCCAAGGTTGAACTCAGTTTATCCGTGCTGCTGGAGGAGAAAGAAAGGACT	820	

Query Match	2.28; .	Score 75;	DB 4;	Length 4403765;
Best Local Similarity	49.04;	Pred. No. 7.6e-08;		
Matches 263;	Conservative 0;	Mismatches 265;	Indels 9;	Gaps 2;
QY	231	CTGGCGCAAAAGGGGTTATAAAGACCTTCTCAAGTGCCTCAAGGTAAATTCCTGCCGCCG	290	
Db	1965431	CTGGACCATCGATGGCGACAAGACACTGCTGGACGCACTCTCGTTGACGCGCGCCCCGG	1965490	
QY	291	GGAGCTGATTTGGCATATGGGCCCTTCAGGGGTGGCAAGTCTACATTCATGAACATCTTT	350	
Db	1965491	TATGCTCTACCGCGCTCATCGTCCGTGGGCGCTGCAAGTCACACTTGTCCCGGTTGGT	1965550	
QY	351	GGCAGGATACAGGGA---GTCTGGAATGAAGGGCGAGATCCTGGTTAATGGAAGGCCACG	407	
Db	1965551	GGCTGGGTATACACACCCCGACGGATGGCAGGTGACGTTTCGAGGGCCACACCTTCACGC	1965610	
QY	408	GGAGCTGAGGACCTTCGCAAGATGTCCTGCTACATCATGCAAGATGACATGCTGTGCC	467	
Db	1965611	CGAATATGCTCTCGCAGCAGGATCGGCATTTGGTCCACAGGACGACGTGGTGCACCG	1965670	
QY	468	GCACCTCACGGTGTGGAGCCATGATGGTCTCTGCTAACCTGAATCTTACTCAGAAATCC	537	

Db	1974962	GCATATCGACACACGGGTGCACAAGCTGTCGGGTGTGTCAACGCAGACGGGGCGCTCGGTGGC	1975021
Qy	642	CTTGGAGCTGGTCAACACACCCGGCTGTCAATGTTCTTTGATGAGCCACCAAGTGGTCTGGA	701
Db	1975022	GCTTGAAGCTTTTTCACCGGGGGCGTCACTGCTGATCCTCTGGACGACCGACATCCGGCGCTAGA	1975081
Qy	702	TAGCGGCTCTTGTGTTCCAAAGTGGTGTGCCCTCATGAAGTCCCTGCACAGGGGGGGCGG	758
Db	1975082	TCCTGCGCTGGACCGGACAGTCTATGACCAATGCTGCGGCACTTGGCCGACACCGCGTGC	1975138

RESULT 4
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-Fls
US-08-232-463-14

[illegible]

```

Db      1254 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1313
QY      3288 ACTCAGAGCATGCTCGCTTTCTCGTGGGTCCAGGCTCATCTGCTTCTGATTCGCCCT 3347
Db      1314 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1373
QY      3348 CCCCCAGGGCTCATTTTCCCCCTTTTTCCTGTACACATCCCTGCTTACCTCCTCTCACCC 3407
Db      1374 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1433
QY      3408 TGCACAGATTTCTC 3422
Db      1434 YGTACCAATCTTC 1448

RESULT 5
US-09-199-637A-111
; Sequence 111, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Ellana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111
; LENGTH: 6909
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-199-637A-111

```

	Query Match	2.0%;	Score 69;	DB 4;	Length 6909;
	Best Local Similarity	47.7%;	Pred. No. 1.4e-07;		
	Matches	244;	Conservative	0;	Mismatches 255; Indels 12; Gaps 1;
Qy	285	CCGCGGGAGCTGATTCGGCATCAATGGGCCCTCAGGGCGTGGCAAGTCTCATATTCATGA	344		
Db					
Db	6049	CAGGGCGAGGTACTGTGTGTGATCGGTCGGTCCGGCAAGTCGACCTTCTCTCGG	6108		
Qy	345	CATCTTGCCAGGATACAGGAGTCTGAAATGAAGGGCAGATCCTGGTTAATGGAAGGCC	404		
Db					
Db	6109	CTGCCTGAACGGCCTGGAGGATTCGACGAAGGCTCGGTGAGCATCAGCGCGTCGACCT	6168		
Qy	405	ACGGGAGCTCAGGACCTTCGCGCAGATGTCCTGCTAC-----ATCATGCAAGA	452		
Db					
Db	6169	GCGCGACCGCAGGACCACATCAATGCTCACCGCCGCGAATCGGCATGTTGTTCCAGCA	6228		
Qy	453	TGACATGCTGCTGCCGCACCTCAGGTGTTGGAGCCATGATGTCCTCTGCTAACCTGAA	512		
Db	6229	TTTTCAACTGTTCCCGACATGACCGTGCTCGAGAACTCTGCGTGGCCCCAACCGCGTGT	6288		
Qy	513	TCATTACTAGAAATCCCGATGTGAAAAACGATCTCGTGACAGAGATCTGACGGCAGCTGGG	572		
Db	6289	GCGCAAGCGCGCAAGGCCGAGCGCAGGCCAAGCGCGGCGCGTGTGTCGCAAGGTCGG	6348		
Qy	573	CCTGATGTCGTCTCCACACAGGACAGCCCTGCTCTCTGGCGGGCAGAGGACGCTCT	632		
Db	6349	CATCGGCGAAGGCGCCAGCAATATTCCTCGCGCTGTGCGGGGGCCAGCAGCAGCGGT	6408		
Qy	633	GGCCATCGCCCTGGAGCTGGTCAACAACCGCCCTGTGATGTTCTTTTCATGAGCCCACCA	692		

Db 6409 GCGGATCGTCGCGCGCTTTGTCATGACACCCCAAGGTGATGCTGTCGACGAACCGACCTC 6468
QY 693 TGTCTGGATAGCGCCTCTTGTTCCTCAAGTGGTGTCCCTCATGACATGCTCCCTGCGCAGGG 752
Db 6469 GCGGCTCGATCCGGAGATGGTGGCGAAGTCTTCGACGTCATGAAGACCTTGCGCGTGA 6528
QY 753 GGCCCGTACCATCATCTGCACCATCCACCCAG 783
Db 6529 AGGCATGACCATGGTCTGCTGACCCACCCAG 6559

RESULT 6
US-08-997-080-88
; Sequence 88, Application US/08997080
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1518 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-997-080-88

Query Match 1.8%; Score 63.2; DB 2; Length 1518;
Best Local Similarity 45.9%; Pred. No. 2.1e-06;
Matches 255; Conservative 0; Mismatches 298; Indels 3; Gaps 1;

QY 281 TCTGCGCGCGGAGCTGATGGCATCATGCGGCCCTCAGGGCTGCGCAAGTCTACATTCA 340
Db 437 TCGCGCGCGGAGTCTTCTCTCATCTCGCGCGCTCGGGTGTGGGAAGACGACACCT 496
QY 341 TGAACATCTTGCAGGATACAGGAGTCTGGATGAAGGGGAGATCCTGGTTAATGAA 400
Db 497 TCGCGATGATCGCGGAGTTCGAGACCCCAAGGCGGATCGCGCTCGAAGGCGCG 556
QY 401 GGCCAGCGGAGCTGAGGACCTTCGCAAGATGCTCTGCTACATCATGTCGAAGATGACATGC 460
Db 557 ACCTGTCGAGGACCCCAACACAGCGCAACGCTCAACAGCGTGTTCAGCACTAGCGCC 616

QY 461 TGTGTCGCGCACCTCAGGTTGGAAAGCCATGATGCTCTCTGCTAACTGAATCTTACTG 520
Db 617 TGTTCGCGCACATGACGGTCTGGGACACAGTCCGCTACGCCCGCCGACGAAGAACTCG 676
QY 521 AGAATCCCGATGTGAANAACGATCTCGTGACAGAGATCTGAGGGACACTGGGCTGATGT 580
Db 677 GCAAGGCGGAGGT--CCGCAAGCGCGTGCAGCGAGCTGCTGGAGATCGTCCGCGTACCG 733
QY 581 CGTGTCTCCACAGGAGGACAGCCCTGCTCTGCGCGGCGAGAGGAGCGCTCTGGCCATCG 640
Db 734 AATTCGCGAGCGAGCGCGCCCGAGCTGTCCGCGCGGAGAGGAGCGGCTGGCTGG 793
QY 641 CCCTGGAGCTGGTCAACAACCCCGCTGTCTATCTTTTGTATGAGCCACAGTGGTCTGG 700
Db 794 CCGGCGCACTGGTGAATACCCAGCGCGCTGCTGCTCATGAACCGCTCGGAGCGCTCG 853
QY 701 ATAGCGCCTCTTGTTCCTCAAGTGGTGTCCCTCATGAAGTCCCTGGGACAGGGGGCCGTA 760
Db 854 ACCTGAAGCTGGCGCCACGTCATGCTGAGCTCAAGCGCATCCAGCGGAGGTCGGGA 913
QY 761 CCATCATCTGCACCATCCACGAGCCAGTCCCAAGCTCTTTGAGATGTTTGACAGCTCT 820
Db 914 TCAGTTCATCTAGTGACCCAGCAGGACGAGGAGCGGCTCAGGATGAGTGACCGCATCG 973
QY 821 ACATCTCTGAGCCAGG 836
Db 974 CGGTGATGAACGCGG 989

RESULT 7
US-08-997-362-88
; Sequence 88, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 88:

Qy	521	AGATCCCGATGTGAA	AAACGATCTCTGTGACAGAGATCCTGACGGCACTGGGCCCTGATGT	580
Db	677	GCAAAAGCGAGGT	---CCGAAAGCGCGTCGACGAGCTGCTGGAGATCTCGCGCTGACCG	733
Qy	581	CGTGCTCCCA	CAGAGACAGCCCTGCTCTCTGCGGGCAGAGAAAGCTCTGGCCATCG	640
Db	734	AATTTGCGAGCG	AGCGCCGCCAGCTGTCCGGGGCAGCAGCGGTGGCTTGG	793
Qy	641	CCCTGGAGCTGGT	CAACAACCGCGCTGTCTATGTTCTTTTGATGAGCCCACTGTCGTCTGG	700
Db	794	CCGGGCACTGGT	GAACTACCCACGCGCTGCTCTCGATGAACCGCTCGAGCGCTCG	853
Qy	701	ATAGCGCTCTGT	TTTCCAAGTGTGTCCCTCATGAAGTCCCTGGCACAGGGGGCCGTA	760
Db	854	ACCTGAAGCTGCG	CCACGTCTATGAGTTCAGAGCTCAAGCGCATCAGCGGAGGTCCGGGA	913
Qy	761	CCATCATCTGCAC	CATCCACAGCCCACTGCAAGCTCTTTGAGATGTTTGACAAAGCTCT	820
Db	914	TCAGTTTCATCT	ACGTGACCCACGACCAAGGAAGGGCGCTCACGATGAGTGACCGCATCG	973
Qy	821	ACATCTTGAGCC	AGGG 836	
Db	974	CGGTGATGAAC	CGCG 989	

RESULT 10
 US-09-324-542-88
 ; Sequence 88, Application US/09324542
 ; Patent No. 6328978
 ; GENERAL INFORMATION:
 ; APPLICANT: Watson, James D.
 ; APPLICANT: Tan, Paul L.J.
 ; APPLICANT: Prestidge, Ross
 ; TITLE OF INVENTION: Methods and Compounds for the Treatment
 ; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
 ; FILE REFERENCE: 11000.1007c1
 ; CURRENT APPLICATION NUMBER: US/09/324,542
 ; CURRENT FILING DATE: 1999-06-02
 ; EARLIER APPLICATION NUMBER: US 08/997,080
 ; EARLIER FILING DATE: 1997-12-23
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: FastSeq for Windows version 3.0
 ; SEQ ID NO 88
 ; LENGTH: 1518
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium vaccae
 US-09-324-542-88

	Query Match	1.8%;	Score 63.2;	DB 4;	Length 1518;
	Best Local Similarity	45.9%;	Pred. No. 2.le-06;		
Matches	Conservative	0;	Mismatches 298;	Indels	Gaps 1;
QY	281	TCTCCGCCGGGAGCTGATTGGCATCATGGGCCCTCAGGGCGTGGCAAGTCTACATTC	A	340	
Db	437	TCGCGCCGGGGAGTTCTTCATGTCTGGCCCGCTCCGGGTCTGGGAAGCACCAAGT	C	496	
QY	341	TGAACATCTTTGGCAGGATACAGGGAGTCTCGAATGAAGGGSCAGATCCTGGTTAAATGSA	A	400	
Db	497	TGCGCATATCGCGGGGATTCAGACCCCGACTGAAGGGCGATCCGGCTCGAAGGCCGC	G	556	
QY	401	GGCCAGGGGAGCTGAGGACCCTTCGGCAAGATGTCTCTACATCATGCAAGATGACATGC	C	460	
Db	557	ACGTGTGAGGACCCCAACCAACAGCGAACGTCACACGGTGTTCCAGCACTACGGC	G	616	
QY	461	TGTCGCCGCACTCACGGTGTGGAAAGCCATGATGGTCTCTGCTTAACCTGAATCTTACTG	G	520	
Db	617	TGTTCCGCACATGACGGTCTGGGACAACGTGCGGTACGGCCCGCAGCAAGAATACTCG	G	676	
QY	521	AGAAATCCCGNTGTGAANAACGATCTCGTGTACAGAGATCTCTGACGGCACTGGGCCCTGATGT	T	580	
Db	677	GCAAAGCGAGGTT---CCGCAAGCGGTGCACAGAGCTGCTGGAGATCGTCCGGTGTACCC	G	733	

QY 581 CGTGTCTCCACACGAGGACAGCCCTGCTCTCTGGCGGCAGAGGAAGCGTCTGGCCATCG 640
Db 734 AATTTGCCGAGCGAGCGCCGCCAGCTGTCCGGCGGCAGCAGCAGCGGTGGCTTGG 793
QY 641 CCCTGGAGCTGCTCAACAACCCGCTGTCTCTTTTGTATGAGCCACAGTGTCTGG 700
Db 794 CCGGGCACTGCTGAACATACCCACGCGCTGCTCTGATGAACCGCTCGAGCGCTCG 853
QY 701 ATAGCGCTCTGTTTCCAAAGTGTGTCCTCATGAAGTCCCTGCGCAGAGGGGCCGTA 760
Db 854 ACCTGAAGCTGGCCACGCTATGAGTTCAGGCTCAAGCGCATCCAGCGGAGGTGGGA 913
QY 761 CCATCATCTGCACCATCCACAGCCAGTGCACAGCTCTTTGAGATGTTTGACAAGCTCT 820
Db 914 TCACGTTTCATCTACGTGACCCACGACAGGAGGCGCTCAGCATGAGTGACCGCATCG 973
QY 821 ACATCCTGAGCGGG 836
Db 974 CGGTGATGAACGCCGG 989

RESULT 11
US-09-205-426-88
; Sequence 88, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 88
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Mycobacterium vaccae
09-205-426-88

Query Match 1.8%; Score 63.2; DB 4; Length 1518;
Best Local Similarity 45.9%; Pred. No. 2,1e-06;
Matches 255; Conservative 0; Mismatches 298; Indels 3; Gaps 1;

QY 281 TCTGCCGCGGAGCTGATGTCATATGCGCCCTCAGGGCTCGCAAGTCTACATTC 340
Db 437 TCGGCCCGGGAGTCTTCTCCATGCTCGGCCCTCGGGGTGTTGGGAAGACGACCGT 496
QY 341 TGAACATCTTGCAGATACAGGAGTCTGGAATCAAGGGCGAGATCTGTTTATGGA 400
Db 497 TCGCGATGATCGGGGATCTGAGACCCCGACTGAAAGGGCGGATCGCGCTCGAAGCGCG 556
QY 401 GSCACGGGAGCTGAGGACCTTCCGCAAGATGCTCTGCTACATCATGCAAGATGACATGC 460
Db 557 AGCTGTGAGGAGCCACCCACACAGCAAGCTCAACAGGTGTTCCAGCACTACGCGC 616
QY 461 TGCTCCGCACTCAGGTGTGGAAGCCATGATGCTCTGCTAACCTGAATCTTACTG 520
Db 617 TGTTCGCGCATGACGCTCTGGGACACGCTGCGCTAGCGGCCCGCGCAGCAGAACTCG 676
QY 521 AGAATCCCGATGTAACACATCTCTGTCACAGATCTCTGACGCACTGGCCCTGATGT 580
Db 677 GCAAGGCGAGGT-----CCGACGCGCTCGAGGAGCTCTGGAGATCTGTCGGCTGACCG 733

QY 581 CGTGTCTCCACACGAGGACAGCCCTGCTCTCTGGCGGCAGAGGAAGCGTCTGGCCATCG 640
Db 734 AATTTGCCGAGCGAGCGCCGCCAGCTGTCCGGCGGCAGCAGCAGCGGTGGCTTGG 793
QY 641 CCCTGGAGCTGCTCAACAACCCGCTGTCTCTTTTGTATGAGCCACAGTGTCTGG 700
Db 794 CCGGGCACTGCTGAACATACCCACGCGCTGCTCTGATGAACCGCTCGAGCGCTCG 853
QY 701 ATAGCGCTCTGTTTCCAAAGTGTGTCCTCATGAAGTCCCTGCGCAGAGGGGCCGTA 760
Db 854 ACCTGAAGCTGGCCACGCTATGAGTTCAGGCTCAAGCGCATCCAGCGGAGGTGGGA 913
QY 761 CCATCATCTGCACCATCCACAGCCAGTGCACAGCTCTTTGAGATGTTTGACAAGCTCT 820
Db 914 TCACGTTTCATCTACGTGACCCACGACAGGAGGCGCTCAGCATGAGTGACCGCATCG 973
QY 821 ACATCCTGAGCGGG 836
Db 974 CGGTGATGAACGCCGG 989

RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 1.8%; Score 62.8; DB 4; Length 4403765;
Best Local Similarity 54.3%; Pred. No. 8.8e-05;
Matches 127; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 549 GACAGATCTGACGGCACTGGGCTGATGTCGTCCTCCACAGAGGACAGCCCTGCT 608
Db 3000173 GATGACGCTGTGGCGCGCTGCGCTGGCCGATGACGCCACACCTAGTGGCAAGTA 3000114
QY 609 CTCTGGCGGCGAGAGAGCGTCTGGCCATCCCTGGAGCTGGTCAACAACCCGCTGT 668
Db 3000113 CTCCAAGGCGATGAGATGCGGCTCACCTTCCCGAGTGCCTGATCAACGACCCGAGCT 3000054
QY 669 CATGTTCTTTGATGAGCCACAGTGTCTGGATAGCGGCTCTTGTTCCTCAAGTGGTGT 728
Db 3000053 GCTGTTCTCGAGAGCCACCTCCGCGGTGAGCCCGGTAACGCCCAAGATCAAGTA 2999994
QY 729 CCTCATGAAGTCCCTGGCAGAGGGGGGGGCTGACATCATCTCCACCATCCACCA 782
Db 2999993 CATCATCTGGACCTGAAGCGCGCGCGCGCATCTTTCTTCTCACCACGACGA 2999940

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:

1149	QY	CTTCTCGTCCATCCTCAGGAGACAGGTCTCTGACCCACCTACGGTTTCATGTCCACAGTGGT	1208
370	Db		429
1209	QY	TATTGGCGTGCATCATCGGCTCTCTACCTGCATATTGGCGAGATGCCAGCAAGTCTT	1268
430	Db		489
1269	QY	CAACAACACGGGTGCTCTTCTTCTCCATGCTGTTCCTCATTTGCGCGCCCTCATPGCC	1328
490	Db		549
1329	QY	AACGTGCTCACCCTTCCGCTTAGAGATGGGGTCTTCATGAGGGAGGACCTCAACTACTG	1388
550	Db		609
1389	QY	GTACAGCCTCAAGCGTATTACCTGGCCAAAGACACATGGCTGAGCTGCCTTTTCAGTGGT	1448
610	Db		669
1449	QY	GTCTCGGSGTGTTACTTGCAGCATTTGTACTGGATGACGGGCGACCGCTGAGACCAG	1508
670	Db		729
1509	QY	CCGCTTCTGCTCTTCTACGCCCTGGCGCACCGGCCACCGGCTTGGTGGCCAAATCTTTGGG	1568
730	Db		789
1569	QY	GTGCTGATCGGAGCTCTCCAACTCCGTACAGGTGGCCACTTTGTGGGCCCAAGTTAC	1628
790	Db		849
1629	QY	CGCCATCCCTGTCTTGTCTTCTCCGGCTCTTTGTGCAGTTTCAAGACCATCCCCACTTA	1688
850	Db		909
1689	QY	CCT 1691	
910	Db		
		CCT 912	

```

RESULT 15
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D9030U1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

	Query Match	1.8%	Score 61.4;	DB 2;	Length 1931;
	Best Local Similarity	43.3%;	Pred. No. 6.5e-06;		
	Matches 287;	Conservative 0;	Mismatches 376;	Indels 0;	Gaps 0;
Qy	1029	CCCTGAGAGACGAGGTCCCTGCCCCATGCCCTCTTGTCTCCGAAAGTGGATCCCAT	1088		
Db	250	CCCTGGTGGCACTTTGCTCTGCTCTGCTGCCCTCCCGTCTCTGCTCTCCCGT	309		
Qy	1089	TGAAGCCACACCTTTGCCACACGACCCCTCACACAGTTCTGCATCCTCTTCAAGAGGAC	1148		
Db	310	CTCTGCTCTCTCCCGGTCTCTGCTCTCTCTCCCGTCTCCCGTCTCCCGTCTCCCGTCTCCG	369		

```
QY 1149 CTTCTGTCCATCTAGGGACAGGCTCTGACCCACCTACGGTTTATGTCCCACTGGT 1208
  ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| |
Db 370 CGTCTCGTCTCTCTCCCGTCCCTCGTCTCTCTCCCGTCTCTCCCGTCTCT 429

QY 1209 TATTGGCGTGTCTATGGCCCTCTCTCTACCTGATATTGGGACGATGCCAGAGGTCTT 1268
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 430 CGTCTCTCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 489

QY 1269 CAACAACACCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1328
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 490 CCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 549

QY 1329 AACTGTGCTACCTTCCCTTTAGAGATGGCGGTCTTCAATAGGGAGCACCTCAACTACTG 1388
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 550 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 609

1389 GTACAGCTCAAGGGTATTACCTGGGCCAAGACCATGGCTGACGTGGCCCTTTCAGGTGGT 1448
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
610 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 669

1449 GTGTCGGTGTCTACTGACATGTCTACTGGATGACGGGCCAGCCCGCTGAGACCAG 1508
  | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 670 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 729

1509 CCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1568
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 730 CCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 789

1569 GCTGTGATCGGAGTGTCTTCACTCCCTACAGGTGGCCACTTTTGTGGGCCCACTTAC 1628
  || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 790 CTTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 849

1629 CGCCATCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1688
  || || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db 850 GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 909

QY 1689 CCT 1691
  |||
Db 910 CCT 912
```

Search completed: June 16, 2003, 13:59:40
Job time : 177 secs


```
Query Match      30.4%; Score 1016.5; DB 22; Length 643;
Best Local Similarity 37.4%; Pred. No. 2.9e-89;
Matches 236; Conservative 103; Mismatches 195; Indels 97; Gaps 11;

QY 52 HLPKRSVDIEFVELSYSVREGFCWKRGRYKTLKLSGKFCRRELIGIMGPSGAKSTF 111
Db 45 YLPSWPAVNLQFSELQYVDPQTNASKT--KTLRRVNGMFHSHeltaIIGPSGAKSTL 102
QY 112 MNILAGYRESGMGOILVNGRPRELTFKMKSCYIMODDMLPHTVLEAMVYSANLNT 171
Db 103 LNLAGEGAVCESGEILVNNSPRDMVRFRMSRYIMOTDVLDSQFTVLEMMILAANKLG 162
QY 172 ENPDVKN--DLVTEITLALGLMSCSHTRTALLSGGQRKRRLAIALELVNPNPVMFFDEPTS 229
Db 163 KELNLQKLEVIDEILGMLKDTLNTMAQKLSGGERKRLCIALELVNPNPVIFLDEPTT 222
QY 230 GLDSASCFQVSLMKSIAOGRITICTIHOPSAKLEPFMDKLYILSGQCIFKGVVNTLI 289
Db 223 GLDDLSSQCIALLKVLAAAGRTVICSIHTPSAKIFEMDLDAVYVLAEGQCVYQCGKSNIV 282
QY 290 PYLKGGLGHCTYHNPADFTIEVASGEYGD--NPMLFRAVQNGLCAMAEKKSPEKNEV 347
Db 283 PELKNEGLCCPIYNPADFTIEVACLEYGGSYHEPMW-EAVNG-----KV 327
QY 348 PAPCPCPPEVDPIESHFTATSTLT-----QFCIL 377
Db 328 ARWSPEA-----VEGETTTTKDTTSGASSFYERQFDEISPKLLEAKCSWMLQYKLL 381
QY 378 FRKTELSILRDTVLTHLRMSHVVGVLIGLLYHLHGDDASKVFNNTGCLFFSMLFMFA 437
Db 382 LVRMLQMRKSYIKLKFYMNIVLALLVGLYF-----415
QY 438 ALMPTVLTFPLEMAVEMREHNLWYSLKAYILAKTMADVPFQVVPVYCSIVVMTGQP 497
Db 416 -----VPEINLLKREYNQWYRLSSYYAAMVTSKLPSMFILAVIYLSIVYLMSSQP 467
QY 498 AETSRELLFSALATATALVAQSLGLLGAASNSLOVATFVGPVTAIPVLFSGFVFSF-- 555
Db 468 LEWFRFAMLTIAFVTALTSDFGLLI--SSRLSLVNAMFMPVLAFLILLSIYGICYGG 526
QY 556 -KTIPTYLOWSSYLSYVRGEGVILTYCMERGD LTC--LEERCPPREPQSTILRALDVE 612
Db 527 GTYISPLMRFLMHLSLRHSMEGLVASLYDYGRADTICDETIFCSFKKSKVLLAFLGFE 586
QY 613 DAKLYMDFLVLGIFFLALRLAYLRLRYVK 643
Db 587 NMHYLWSLSCLMSFYLLFTVAAYFMIRLRK 617
```

Search completed: June 10, 2003, 16:44:32
Time : 76 secs

QY 145 YIMODMLPHTLVLEMMVYANLNL--TENPDVKNLDVTEITLTAIGLMSCSHRTALLS 202
 Db 215 YIHQDRLRPQLLGEIMLAHLKFKYKMDLIKHLISLLCLDHRYNVPTGKLS 274
 QY 203 GGOKRLAIALELVNPNPVMFEPTSGLDASCFQVWSLMKSLAOGGRTIICITHQPSA 262
 Db 275 GGOKRLAIALELVNPNPVMFEPTSGLDASCFQVWSLMKSLAOGGRTIICITHQPSA 334
 QY 263 KLFEMFDKLYLSQOGCIFKGVVNTLIPYLKGLGLHCPHTVNPADFIEVASGEY-GDLN 321
 Db 335 LIFEMFDKLYVVDGCHMYGPPVRELVPFLADQOLVCPSHVNPADYLLVAVAGEHQDLN 394
 QY 322 PMLFRA-----VQGLCMAAEK----- 338
 Db 395 ELIHAANKYIEDVDRHRYMSSDDMARLIVESIKENMGKAVVKTSEALAAFAAQFSFD 454
 QY 339 --KSSP-----EKNEVPAPCPPC-----PPEVDPIESHTF 366
 Db 455 YKPSFOELALEIKALSGGPESADPDLEKRLRPOPOPLAKAGELARPP--NAIRASAP 512
 QY 367 ATSTLTQFCILFKRTFLSILRDVLTHLRPMHSHWVIGVLGLLGLYHLHGDDASKVFNNTGC 426
 Db 513 ----LMQYVLLMQRLICAKRNIFELLARIFSHIFGVGVGYLYMNVGNNAQSVLGNVY 568
 QY 427 LFFSMFLFPAALMPVLPFLPMEAVMFMREHLNYSKAYIAKTMADVPFQVCPVY 486
 Db 569 LYGSTLLVYTKMAVVLTPLEIDMLTREHFNRYKLGPFYLSLISFEIPFQSLCTAY 628
 QY 487 CSTVYMTGPAETS-RFLFSLATATATAVAQSLGILLIGAAASNSLOVATFVGPVTAIPV 545
 Db 629 IISVHLTGNDATDSFRIFYFMLLGLMASLSAQANGFFVGTATLPT-KLAVFLGPILAVMF 687
 QY 546 LIFSFGFVSKTIPTIYLOWSSYLSYVRYGEGVILTIYGMERGDLTCLEER--CPTREPQ 603
 Db 688 SVF-GFCTRYDITPLFRMMHLSYFRAGPHGALNAYGMDRPFLECPETAMYCHFRSPK 746
 QY 604 SILRALDVEDAKYMLDFVLGIFFLALRLLAYLVLYRV 642
 Db 747 VFLKYMISDVHMSDCLILMGIVIGVMHVLTLITLWHKL 785
 RESULT 13
 ID ABB60914 standard; Protein; 643 AA.
 AC ABB60914;
 XX ABB60914;
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 9534.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 XX W0200171042-A2.
 PN 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL05017.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions
 XX Disclosure; SEQ ID NO 9534; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 643 AA;

Query Match 30.4%; Score 1016.5; DB 22; Length 643;
 Best Local Similarity 37.4%; Pred. No. 2.9e-89;
 Matches 236; Conservative 103; Mismatches 195; Indels 97; Gaps 11;
 QY 52 HLPKRSADVIEFVELSYSVREGPCWKRKYKTKLLKCLSGKFCRRELIGIMPAGAGKSTF 111
 Db 45 YLPSWPAVNLQFSELQVVPDQTNASKT--KTLRRVNGMFMHSHELTAITIGPSGAKGTTL 102
 QY 112 MNTLACYRESGKGQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTIVLEAMVYSANLNT 171
 Db 103 LNLLAGFGAVCESGEILVNNSPRDMRFRKMSRYIMQDVLDSQFTVLEMMILAAANLKL 162
 QY 172 ENPDVKN--DLVTEILTALGLMSCSHRTALLSGGOKRLAIALELVNPNPVMFFDEPTS 229
 Db 163 KELNLQKLEVIDEILGMLRLKDTLNTMAQLSGGERKRLCIALELVNPNPVLDEPTT 222
 QY 230 GLDSACQFVWSLMKSLAOGGRTIICITHQPSAKLFEMFDKLYLSQOGCIFKGVVNTLI 289
 Db 223 GLDDLSSQCIALLKVLAAAGRTVIGSIHTPSAKIFEMDLDAVYVLAEGQCVCYCKGKSNIV 282
 QY 290 PYLKGILGHCTYHNPADFIIEVASGEYGD--NPMLFRAVQNGLCMAEKKSSPEKNEV 347
 Db 283 PFLKNFLGCCPTYNPADFIIEVACLEYGSGYHEPWV-EAVNG-----KV 327
 QY 348 PAPCPPEPPEVDPIESHTATSTLT-----DDEISPKLEAKCSNWLQYKLL 377
 Db 328 ARWSPA-----VEGETTTTKDTTSGASSFYERQDFDEISPKLEAKCSNWLQYKLL 381
 QY 378 FKRTFLSILRDTVLTHLRFMSHVIVGLIGLLYHLHGDDASKVFNNTGCLFFSMLFLMFA 437
 Db 382 LVRMLIQMWRDKSYIKLFYMNIVLALLVGLSYF----- 415
 QY 438 ALMPTVLTPLPMEAVMFMREHLNYSKAYIAKTMADVPFQVCPVYCVSIYVWMTGQP 497
 Db 416 -----VPTENLLKREYNQWYRLSSYYAAMVTSKLPMSFILAIVLSIVLWSSQP 467
 QY 498 AETSRELLFSALATATVAQSLGILLIGAAASNSLOVATFVGPVTAIPVLFSGFFYSF-- 555
 Db 468 LEWFREAMLFTIAFVTALTSDSFGLLI--SSRLSLVNAMMGPVLAIVPLLSIYIGYGG 536
 QY 556 -KTIPTYLQWSSYLSVRYGFEVILTIYGMERGDLTC--LEERCFFREPQSIILRALDVE 612
 Db 527 GTYISPLMRFLMHLSYLRHSMGLVASLDYDGRADTICEDTEIFCSFKKSKVLLAFLGFE 586
 QY 613 DAKLYMDFLVIGIFFLALRLAYLVLYRVK 643
 Db 587 NMHYLWSLSCLMSFYLLFTVAAYFMIRLUK 617

RESULT 14
 ABB66977
 ID ABB66977 standard; Protein; 643 AA.
 XX ABB66977;
 AC ABB66977;

PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL08827.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

Disclosure; SEQ ID NO 20964; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).
The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO.
at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 615 AA;
SQ

Query Match 31.9%; Score 1067; DB 22; Length 615;
Best Local Similarity 36.5%; Pred. No. 3.5e-94;
Matches 224; Conservative 121; Mismatches 229; Indels 40; Gaps 9;

QY 59 VDEFEVLSVSVREGPCWKRKYKTLKLSGKFCRRELIGIMPSGAGKSTFMTILACY 118
DB 13 IDHFDLVQVNP---KKPKSVLKGIGTKFSGELTAIMGPSGAGKSSLMNLTGL 69
QY 119 RESGKMGQILVNGRPRRLTRFMKSCYIMQDMLPLHVLVLEAMVYSAANLITE---NPD 175
DB 70 TRSGVSGKI-----EICKARKLCGYIMQDDHFFPYFTVEETMLAATLKISNQCVSLK 122

176 VKNDLVTEILTALGLMSCSHRTALSGGQRRLAIALELVNPPVWFDEPTSGLDSSAS 235
DB 123 EKRTLIDYLLNSLKLTKTRQTKCSNLSGGKKRLSIALELIDNPVLFLEPTTGLDSSS 182
QY 236 CFQVWSLMSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVTNLIPYLKGL 295
DB 183 SFTIQLLGLRANEGRTIVCTHQPSTNYINFLNVLVLSAGRCTYQGTPTQFMFLSSV 242
QY 296 GLHCTPHNPADFIIEVASGEXGDLNPLFRAVQN-----GLCMAEKKSPEK 344
DB 243 GLECPYPHNPADFLLECANGDYQDQTEALAAEAKDIRWYDQQLMQGEDADAPSETQVAK 302
QY 345 -NEVPAPCPPEVDPIE-----SHTFATSTLTQFCILKRTFLSLTRDTVLT 392
DB 303 FNESQSP-GQVQVQVQKIEIQNMESKDLTKHTPYPTMYRLWLLIGRCHLOFFRDWTLT 361
QY 393 HLRFMSHVIGVLIGLILYHLHGDSKASKVFNNTGCLFFSMLFLMFAALMPTVLFPLEMAV 452
DB 362 YLKLGHVLCISLIGLFGDGSNATKQISNMGIMHCVLYWYTIMPGILRYPAIEI 421
QY 453 FMREHLNMYSLKAYLAKTADVPVQVVCVIVYMTQQAETSRFLFSALATA 512
DB 422 IRKETENNYKLRYYLATIITSTPVHIIIFSTVYITIGYILMTDQPVEMDFRYKLLSAVV 481
QY 513 TALVAQSLGILLGAASNSLOVATFVGPVTAIPVLLSFGFVSPKTTPTYLQSSVLSYVR 572
DB 482 VTICADGLGVFLGLNPNYN-GTFVGAVSTSCMLMFSGFLILNLPAAAMRFMAYISPLR 540

QY 573 YGPEGVILTYGMERGLDTC--LEERCPPREFPOSILRALDVEDAKLYMDFLVLCIFELAL 630
DB 541 YALENMVISLYGNORGLICPPTFEYCHFKNAVTVLRQGMEDGDFGHNLMILIQIAIF 600
QY 631 RLLAYLVLRVYKVS 644
DB 601 KVLSTYFTLKHKIKT 614

RESULT 12

ABB67949
ID ABB67949 standard; Protein; 788 AA.

XX AC ABB67949;

XX 26-MAR-2002 (first entry)

DT DE Drosophila melanogaster polypeptide SEQ ID NO 30639.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.
N-PSDB; ABL12052.

New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

Disclosure; SEQ ID NO 30639; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 788 AA;

Query Match 31.4%; Score 1051; DB 22; Length 788;

Best Local Similarity 36.2%; Pred. No. 1.8e-92;

Matches 233; Conservative 103; Mismatches 226; Indels 117; Gaps 16;

QY 50 FSHLPKRSAVDIEFVELSVSREGPC-WRKRGYKTLKLSGKFCRRELIGIMPSGAGK 108

DB 98 FDHC---DPVDIQADVRYTVKKFSFPERKFVTKELHGLNGSFRSGELTAIMGPSGAGK 154

QY 109 STFMNITLAG-----FRESCKGQILVNGRPR--ELRTFRKMSC 144

DB 155 STLLNVMGFCNGRGKDDAPGTRVFIARVRYSTGVSGDIRNRPMPASSRFRFQMLC 214

QY 337 EKK-----SSP-EKNEYPA-----PCPP----- 353
 Db 446 -KKVDQLMAAGLITPVTAPVMTSTVPAHFMQGNTEKPLTPINELSSRVWDSTAGIGNAG 504
 QY 354 --CPP-----EVDPIESH-----TFATSLTQFCILFKRTFSLRDTVLTHLR 395
 Db 505 SCCPKKKKKNSRPAIELOP--SHLCKRONIYATPFYRQSLILLVTFELLINRDSLTMMR 562
 QY 396 FMHSHVIGVLIGLXHLIGDDASKVFNNTGCLFFSMLFMAALMPTVLTFFLENAMVNR 455
 Db 563 FAHLITGLITGLTFYFGINDAAQTLNIFRYLFYIMFYCAFSGILVKFPLEPIYSR 622
 QY 456 EHLNWTYSKAYIAKTMADVPQVPCVYCSIVYVMTGQPAETSRFLFSALATATAT 515
 Db 623 EHENRWSLYRAYIVAITLADLPITQICSALFIVPTVLTQOPLKRWFCMFFLIVFTAL 682
 QY 516 VAQSLGLLGAANSLSIOVATFVGPVTAIPVLLFSGFFVSFKTIPTIYLOWSSLSVRYCF 575
 Db 683 VQSIGLAVGAAL-SLKLSIILGPFYCFQFSGFFLMEKDPVFLRMDFISFLKYSL 741
 QY 576 EGVLTITGMERGDLTCLTEERCPFPREPQSIILRALDVEDAKLYMDFLVLGIFFLALRLLAY 635
 Db 742 EGATWAIQFYDREPLACNELYCHLRHPQILASLDWANGNYTLALIFLFGVLVFLILAF 801
 QY 636 LVLRYRVK 643
 Db 802 YIMSFRLR 809

RESULT 10

AB64566
 ID AB64566 standard; Protein: 609 AA.

XX AC AB64566;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 20490.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PP 23-MAR-2000; 2000US-191637P.

XX QY 11-JUL-2000; 2000US-0614150.

XX PA (PERE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL08669.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions

XX PS Disclosure; SEQ ID NO 20490; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 609 AA:

Query Match 32.5%; Score 1086.5; DB 22; Length 609;
 Best Local Similarity 36.9%; Pred. No. 4.5e-96;
 Matches 238; Conservative 118; Mismatches 234; Indels 55; Gaps 10;

QY 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAVD 60
 Db 1 MADNAVOAQPNGLGP-----QKQKALE 22

QY 61 IFEVELSVSRGPGCMKRKRGYKTLKCLSGKFCRRELIGMGSPGAGKSTFNNILAGYRE 120
 Db 23 LHFQSVSYSLKGA---TKGSTPIINACGVKSGRLTAILGSPGAGKSTLNALAGFKL 78

QY 121 SGMKGQILVNGRPRELRTFRKMSCYIMQDMLLPHLTVLEAMVMVANLNLN---PDVKN 178
 Db 79 QGVTFQFLNGRPRDIMSRKMSAYIAQNFVMLNLTVEETLRVSTDLMKPSSTAAQEQ 138

QY 179 DLVTEILFALGMLSCSHRTALLSGQQRKRLAIALELVNPNVPMFFDEPTSGDLSASCF 238
 Db 139 KIIDDIIILQSCRRRTLVKNLSGGEHKKRLSIGIELVTNPIMPFDPTSGDLCVGSYQ 198

QY 239 VVSLMKSLAOGGRTIITIHQPSAKLFEMDFKLYLTLSOGCIFKGVVTVNLIPYLKGLGLH 298
 Db 199 VICHQRLAHGRIVVVVHQPSRLQFLDDVVLVLAHGEVLYAGEQRMLPTFAQSGHI 258

QY 299 CPTYHNPADFIIEVASGEYGDNLNPLFRAVQNGLCAMAEKSSPEKNEVPAPCPPEPV 358
 Db 259 CPQYNPADFALEVCS-QSSTTERCESLITQNM--MNSTASNVMVKLOVDEETALIDVHK 315

QY 359 DPDE-SHTFATSTL---TQFCILFKRTFSLRDTVLTHLRPMHSHVVGVLGLLYHTG 414
 Db 316 DALDLHLRGKEQGVFWTQLSVLLRRHLRSMRDMFAVQMRMLVMHVVALLLGVVYWGIG 375

QY 415 DDASKVFNNTGCLFFSMLFMAALMPTVLTFFLEMAVEMREHLNAYWSLKAYILAKTWA 474
 Db 376 GDAQIVSNVSCLFVILFVFNAGNAPSIILCMQDSAVFIREYNGWISLGAYIUSKVLA 435

QY 475 DVPFQVCPVYCSIVYVMTGQPAETSRFLFSALATATATVAQSLGLITGAASNLQVA 534
 Db 436 DLPLQLTCTPTMFIISIGYFMTGQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 494

QY 535 TFGVPTAIPVLLFSGFFVSFKTIPTIYLOWSSYLSYVRYGFEGLVITTYGMERGDLTC-- 592
 Db 495 IFLVPSATIPFLFSGFFIRLNLNLSWFLRPICDVSSFFRYFEGLMRAIYGYDRGECEYA 554

QY 593 LEERCPPFPQSIILRALDVEDAKLYMDFLVLGIFFLALRLLAYLV 637
 Db 555 TSNFCYYRTAEQFLKDFQMEGNEFDMDMAVLGI-FLILLLLAFFV 598

RESULT 11

AB64724
 ID AB64724 standard; Protein: 615 AA.

XX AC AB64724;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 20964.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

QY 542 AIPVLLFSGFFSVFKTIPTIYLOWSSYLVYRGFEGVITTYGMERDGLTCLLEPCPFRE 601
ABB61867
Db 585 SVFLLFSGFFSVFNAIPVYLRWITVSYRIGFECTALATYSFNRKTLQCHAPYCHEFY 644
QY 602 POSILRALQVEDAKLYMDFLVLGIFLALRLAYLVLYRVKSER 646
Db 645 PQTLEDLQEDADFDALDIAALCLIFVLLRIFAFLEFLRWKLKSTR 689

RESULT 7

ID ABB61867 standard; Protein; 689 AA.

XX AC ABB61867;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 12393.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL05970.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

XX PS Disclosure; SEQ ID NO 12393; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).XX CC The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 689 AA;

Query Match 39.5%; Score 1322; DB 22; Length 689;

Best Local Similarity 43.4%; Pred. No. 8.3e-119;

Matches 275; Conservative 126; Mismatches 215; Indels 18; Gaps 10;

QY 22 VTLEDGAEF---PVLTHLKKVNHTEAQRFSHLPRKSAVDIEFVELSVYREGPCWRK 78
Db 65 VNTNNEFSKQPLPEVVEEVEHFDIDA--LNNLPAREPVDMEFVELSLTVKLG---FN 119QY 79 RGYKTLKLCISGRFCRRELIGMPSGAGKSTPMNITAGYRESGMKGQILVNGRPRELRT 138
Db 120 RGSKEILHNVCGRFPCSQIAIAMPSPGACKSTLLDALSGETTGVDSILLNGRRRLPS 179

QY 139 FRKMSYIMQDDMLLPHLVLEAMWVSANLNL--TENPDVKNDLVTEILTALGLMSCSHT 196

Db 180 FRMSCYITQDDRLQPLLTAVNNMHIAADLKLCQTVSVEKESRIEDILLGLYNHDTQ 239
QY 197 RYALLSGGOKRLAIALELVNPNPVMFDEPTSGLDASCFQVVSIMKSLAOGRTIIC 256
Db 240 LTMRLSGGOKRLUSIAMLINNPTVMFLDEPTTGLDSSCTKVLLELLKLTSGCRTIIC 299
QY 257 IHOPSAKLFEMFDKLYLSQOGCIFKGVVTNLTIPYKLGHLGCHPTYHNPADFIEVASGE 316
Db 300 IHOPTAKLFQIFEDQVVLVSAGNCVYOGSTQKLVPLQSVLDLPCPMYHNPADYIELACGE 359
QY 317 YG-DLNPMLFRAVQNGLC--AMAEKSSPEKNEVPAPCPPEPPVDPIESHFPA--TSTLTQ 373
Db 360 YGYDKVDTLTKLATENGSCLTWFHNPASAVLRAEYLMRKYPIPKT---KSRLEDTSYSNQ 416
QY 374 FCILPKRTFLSLRLDVTLTHLRFMSHWIGVLTLGLLYLHIGDDASKVENNTGCLFFSMLF 433
Db 417 CSVLLRRGFIAKARDTMTHLRIGVNIAAVAFGAMYDHTGREGSVLDNYNLLFAILMH 476
QY 434 LMFAALMPTVLTFFLEMAVFMREHLNYSKAYYLAKTMADVPFQVGVVVCISIVYWM 493
Db 477 HSMTTMMLTVLTFFPMDISILIKEHFNRYSLKAYYTAMTLVDLPISIISCFEFTVIVYLW 536
QY 494 TQOPATSRFLPSALATATATVAQSLGLLGAANSLOVATFVGVPTAIPVLLFSGFV 553
Db 537 SYQPMEWIRFEFFSISLLTVFVGHSGFMIGAWFDVN--GTFLAPVLTPMMFAGFV 595
QY 554 SFKTIPTYLOWSSYLVYRGFEGVITTYGMERDGLTCLLE-ERCPEPOSTLRALDVE 612
Db 596 TLRDPSYLKWSHISYLRIGLEGFISAIYGLDRGLACEEAPYCHYRYPKKELETWIR 655
QY 613 DAKLYMDFVLVGIFFLALRLAYLVLYRVKSER 646
Db 656 GDOFNDVIALGVMLVFRFVSIVVLKAKIKSIR 689

RESULT 8

ABB59648

ID ABB59648 standard; Protein; 620 AA.

XX AC ABB59648;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 5736.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03751.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

XX PS Disclosure; SEQ ID NO 5736; 21pp + Sequence Listing; English.

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 832 AA;

Query Match 51.5%; Score 1723; DB 22; Length 832;

Best Local Similarity 47.1%; Pred. No. 1.6e-157;

Matches 351; Conservative 109; Mismatches 153; Indels 132; Gaps 12;

QY 10 GCGLGCAVAMVLTEDGAPPPVLTTHLKKVENHTEAQR-----PSHLKRSADVIEFV 64

DB 104 GLGNNHALAPKVANNSSGGSP-----NGOKKGTIALSHLPORPPVDIEFC 148

QY 65 ELSYSVREGPCWRKRGYKTLKLSGKFRRELIGIMPSSGAGKSTFMNLAGYREGMK 124

DB 149 DISYSVTDS---HRRGFKTLKSVGKFRNGEITAIMGPSGAKSTLMNLAGYKKAQLS 205

QY 125 GOILVNGRPRELRTFRKMSCYIMODDMLPHLTVLEAMVYSANLNTEN--PDVKNDLVT 182

DB 206 GSVLINSKERNLRFRKRLSCYIMODDVLIANLTVREAMVVAANLKLGNMISYAKVVVE 265

QY 183 EILTALGLMSCHTRTALLSGGORKRLAIALELVNPPVMEFDEPTSGLDSASCFQVVS 242

DB 266 EILLETGLKESVNTLCNLSSGORKRLSIALELVNPPVMEFDEPTSGLDSSTCFQLIS 325

QY 243 MKSLAOGGRTIICITIHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLPIYKLGGLGHCPY 302

DB 326 LRSIARGRTIVCTIHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLPIYKLGGLGHCPY 385

QY 303 HNPADFTIEVASGEYGDNLNPLFRAVONGLC-----FATSTLTQFCILF 378

DB 386 HNPADYVLEVASGEYGDVAPKLVDAVSKGACKKYAHKDYVLTILAQKGNNDIIGSGGA 445

QY 334 --AWA-----EKKSSPEKNEVPAPCP--PPEVDP----- 360

DB 446 ENAMAILTEDEKPPLEDROLEPSIPVDDPAELKPKLETQQSONSDCVNNMPTNAVDD 505

QY 361 -----IESHT-----FATSTLTQFCILF 378

DB 506 SCFSFSSKSGQNAVGGSGGSAVVCMTSLDSDSHSVVTLPNKGTGFTSGWTQFWILL 565

QY 379 KRFLSLTRVTLTHLRFMSHVIGVLGLYHIGDDASKVFNNTGCLFFSMLFMFAA 438

DB 566 KRSEFTILRDKMLTHMLRFLSHVTVGAIIGMIYVDVGNFASKIMSNAGCIPFVSLFTFTA 625

QY 439 LMPVLTFFLEMAVEMREHNLNYSKAYVLAKTADVPQVYCPVYVCSIVYVMTGCPA 498

DB 626 MMTILTFPTMSESVFVREHNLNYSKAYVLAKTADVPQVYCPVYVCSIVYVMTGCPA 685

QY 499 ETSRFLFSALATATALVAQSLGILLGAASNSLQVATFVGPVTAIPVLLFSGFVSKTI 558

DB 686 ELERSVSMFVLCVNSLVAGSLGILLGAGMN-IETGVFLGPVTTIPTILFSGFVNFDTI 744

QY 559 PTYLOWSSLYSVRYGFEVGLTIYGMERGLTCLERCPCFPQSIIRALDVEDAKLYM 618

DB 745 PGYLOWTVYVYRYGFEVGLTIYGMERGLTCLERCPCFPQSIIRALDVEDAKLYM 804

QY 619 DFIVL-GIFLALRLAYLVLRV 642

DB 805 DAVALLGIGFF-AURITAYFLRWKL 828

RESULT 6

AA778980

ID AAY78980 standard; Protein: 689 AA.

XX

AC AAY78980;

XX 05-JUN-2000 (first entry)

XX Silkorm Bm white 2 amino acid sequence.

XX Bm white 2; silkorm; visual characteristic; eye colour; egg colour.

XX

OS Bombyx mori.

XX JP2000000094-A.

XX 07-JAN-2000.

XX 15-JUN-1998; 98JP-0166944.

XX 15-JUN-1998; 98JP-0166944.

XX (NORQ) NORINSUISANSHO SANSHI KONCHU.

XX WPI: 2000-199543/18.

XX N-PSDB; AA298624.

XX Silkorm Bm white gene - for changing visual characters such as egg

XX colour and eye colour of silkorm

XX Disclosure: Page 11-13; 16pp; Japanese.

XX This sequence represents the silkorm Bm white 2 amino acid sequence. The

XX Bm white 1, 2, and 3 genes are identified in the invention. Bm white 1

XX and 2 are homologous to the white gene of humans and mice, and Bm white 3

XX is homologous to the white gene of the mosquito and fly. The genes can be

XX used for changing visual characteristics such as egg colour and eye

XX colour of the silkorm.

XX Sequence 689 AA;

Query Match 45.2%; Score 1513; DB 21; Length 689;

Best Local Similarity 45.2%; Pred. No. 2.6e-137;

Matches 319; Conservative 97; Mismatches 193; Indels 96; Gaps 8;

QY 13 LGGCAVAMVLTEDGAPPPVLTTHLKKVENHTEAQRFSHLKRSADVIEFVLSVRE 72

DB 10 IGKSGVKVTT-----PTQHKTLHLKRPQVDFLAFTDLYKVKQK 49

QY 73 GPCWRKRGYKTLKLSGKFRRELIGIMPSSGAGKSTFMNLAGYRESCKQLVNGR 132

DB 50 G-----RKSNNKTLNLSVSGRLSGELTAILGPSGAGKSTILNLTGNTSGMEGSIIVNGM 106

QY 133 PRELRTFRKMSCYIMODDMLPHLTVLEAMVYSANLNTENP--DYKNDILVTILTLALGL 190

DB 107 ERNLSFRKLSYIMODDMLPHLTVLEAMVYSANLNTENP--DYKNDILVTILTLALGL 166

QY 191 MSCSHTRTALLSGGORKRLAIALELVNPPVMEFDEPTSGLDSASCFQVYVLSMKSQAQG 250

DB 167 MEHRKMTSNLSSGQKRLSIALELVNPPVMEFDEPTSGLDSASCFQVYVLSMKSQAQG 226

QY 251 RTIICITHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLPIYKLGGLGHCPYHNPADFTI 310

DB 227 RTIICITHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLPIYKLGGLGHCPYHNPADFTI 286

QY 311 EVASGEYGDNLNPLFRAVONG-----LC 333

DB 287 EVSCEYGNNTGKLVRAIDNGNNDIRNGMPFPDQTPDYNKDKMEVSLRAGSNWKNLDS 346

QY 334 AMAEKSSPEKN-----EVPAPCPCEPVEDPTES----- 363

DB 347 QVOKEKFDNNSGANGHGNLQILOYATSEV-SKGEPLSLQMDTKQDNAAAALLGTEG 405

QY 364 --HTEATSTLTQFCILFRTFLSILRDTVTLHRFMSHVIGVLGLYHIGDDASKYFV 421

DB 406 SPERYATSEWKQFVVLKRTLLDSRRDWTLMYURLFAHLVGLVGLYHIGDDASKYFV 465

QY 422 NNTGCLFFSMLFMFAALMPTVLTFFLEMAVEMREHNLNYSKAYVLAKTADVPQVYV 481

DB 466 SNLGLFFENMLFMTSMITILSPFLPVLVKEHFNWYSLRSYLAITSIDIFPQAI 525

QY 482 CPVYCSIVYVMTGQPAETSRFLFSALATATATVAQSLGILLGAASNSLQVATFVGPVY 541

DB 526 FCIIYVIVVILTSQPPVWFRFAMFLSSCILLISFVAQSVGLVVGAAAMN-VQNGVFLAPVM 584

KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
 KW immunosuppressive; gene therapy; cytokine cell proliferation;
 KW cell differentiation modulator; immune disorder; infection; cancer;
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.

OS Homo sapiens.

XX WO200153453-A2.

XX 26-JUL-2001.

XX 23-DEC-2000; 2000WO-US34960.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-052317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0820312.

XX 03-AUG-2000; 2000US-0853450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 30-NOV-2000; 2000US-0250583.

XX (HYSE-) HYSEQ INC.

PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
 PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Drmanac RT;

DR WPI: 2001-488707/53.

DR N-PSDB: AAH90113.

XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful
 for treating e.g. cancer and immune deficiency disorders.
 XX Claim 10; Page 639-640; 648pp; English.
 XX The present sequence is one of 251 novel human polypeptides encoded
 by a bone marrow-expressed polynucleotide. The polynucleotide and the
 polypeptide encoded by it are useful in the treatment of various
 CC immune deficiencies and disorders. The deficiencies and disorders may
 CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
 CC infection, or may result from an autoimmune disorder, a coagulation
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
 CC suppression of an inflammatory response or treatment of a nervous
 CC system disorder such as Alzheimer's disease. Detection of the presence
 CC or increased expression of the polynucleotide or the protein it
 CC encodes is useful for the diagnosis and/or prognosis of one
 CC or more types of cancer. The polynucleotide and polypeptide can be
 CC used as nutritional sources or supplements and in the screening of
 CC chemical compounds as potential drugs.

XX Sequence 935 AA;

Query Match 68.7%; Score 2299; DB 22; Length 935;

Best Local Similarity 81.3%; Pred. No. 3e-213;

Matches 457; Conservative 3; Mismatches 4; Indels 98; Gaps 2;

QY 80 GYTKLCLSGKRCRELIGIMGPGAGKSTFNILAGYRESGKMGQILVNGRPRELRTF 139

DB 19 GYTKLCLSGKRCRELIGIMGPGAGKSTFNILAGYRESGKMGQILVNGRPRELRTF 78

QY 140 RKMSCIYIMQDDMLPLHITVLEAMVMSANLMTENPDVKNDLVTILTALGLMSCSHRTA 199

DB 79 RKMSCIYIMQDDMLPLHITVLEAMVMSANLMTENPDVKNDLVTILTALGLMSCSHRTA 138

QY 200 LLSGQQRKRAIALELVNPNPVMFFDEPTSGLSASCFQVYSLMKSIAQGGRTICTIHQ 259

DB 139 LLSGQQRKRAIALELVNPNPVMFFDEPTSGLSASCFQVYSLMKSIAQGGRTICTIHQ 198

QY 260 PSKALFEMDKLYLSGQOCIFKGVNTNLPYLKGLCHCPTVHNPAHFIEVASGEYGD 319

DB 199 PSKALFEMDK-----CIFKGVNTNLPYLKGLCHCPTVHNPAHFIEVASGEYGD 250

QY 320 LNPMFLRAVONGLCAMAEKSSPEKNEVPAPCPPEVDPVIESHTFATSTLTQFCILFK 379
 DB 251 LNPMFLRAVONGLCAMAEKSSPEKNEVPAPCPPEVDPVIESHTFATSTLTQFCILFK 310
 QY 380 RTFSLSLRDTVTLTHLRFMSHVVGVLIGLLYLHIGDDASKVFNNTGCLFFSMLFLMFAAL 439
 DB 311 RTFSLSLRDT----- 320
 QY 440 MPTVLTFFPLEMAVFMREHLNYSKAYLAKTMAVDFVQVCPVYVCSIVYVMTGQPAE 499
 DB 321 -----WCVPVYVCSIVYVMTGQPAE 340
 QY 500 TSRELLFSALATATATVAQSLGLLGAASNSLOVATFVGPTAIPVLLFSGFFVSEKTIIP 559
 DB 341 TSRELLFSALATATATVAQSLGLLGAASNSLOVATFVGPTAIPVLLFSGFFVSEKTIIP 400
 QY 560 TYLQWSSYLSYVRYGFEVILTYIGMERGDLTCLERCPCPREPOSTILRALDVEDAKLYMD 619
 DB 401 TYLQWSSYLSYVRYGFEVILTYIGMERGDLTCLERCPCPREPOSTILRALDVEDAKLYMD 460
 QY 620 FLVLGIFFLALRLAYLVLERYR 641
 DB 461 FLVLGIFFLALRLAYLVLERYR 482

RESULT 5

ABB59544

ID ABB59544 standard; Protein; 832 AA.

XX AC ABB59544;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 5424.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX DR N-PSDB; ABL03647.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

XX PS Disclosure; SEQ ID NO 5424; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB57737-AB572072).

XX CC The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO

PI Patterson C, Gandhi AR, Hafalia AJA, Tribouley CM, Wallia NK;
PI Au-Young J, Walsh RT, Ramkumar J, Lu Y, Lu DAM, Azimzai Y, Lal P;
PI Elliott VS, Nguyen DB, Xu Y, Seilhamer JJ, Borowsky ML, Khan FA;
XX Kearney L, Thangavelu K, Das D, Policky JL;
DR WPI: 2002-205969/26.
DR N-PSDB: AAL44669.
XX
XX New human transporters and ion channel polypeptides for diagnosing,
PT treating or preventing transport, neurological, muscle, immunological
PT and cell proliferative disorders
XX
XX Claim 1; Page 146-147; 230pp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human transporter and ion channel proteins, designated
CC TRICH-1-TRICH-32. The sequences can be used in the treatment of
CC transport, neurological, muscle, immunological and cell proliferative
CC disorders. The present sequence is a protein of the invention.
XX
XX Sequence 646 AA;
SQ
Query Match 99.88; Score 3339; DB 23; Length 646;
Best Local Similarity 99.78; Pred. No. 3.2e-314;
Matches 644; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MAEKALEAVGCGLPGAVANAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAVD 60
Db 1 MAEKALEAVGCGLPGAVANAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAVD 60
Qy 61 IEFVELSYVREGPCWKRKGYKTLKCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRE 120
Db 61 IEFVELSYVREGPCWKRKGYKTLKCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRE 120
Qy 121 SGMKGQLVNGRPRLTFRKMSCTIMODDMLLPHLTVLEAMVSNLNTENPDVKNDL 180
Db 121 SGMKGQLVNGRPRLTFRKMSCTIMODDMLLPHLTVLEAMVSNLNTENPDVKNDL 180
Qy 181 VTEILTALGLMSCSHTTALLSGQKRRLAIALELVNPPVPMFDEPTSGLDSASCFQV 240
Db 181 VTEILTALGLMSCSHTTALLSGQKRRLAIALELVNPPVPMFDEPTSGLDSASCFQV 240
Qy 241 SLKSLAQGGRTICTTHQPSAKLFEMFDKLYLSQGCIFKGVVTVNLIPYLKGLGLHCP 300
Db 241 SLKSLAQGGRTICTTHQPSAKLFEMFDKLYLSQGCIFKGVVTVNLIPYLKGLGLHCP 300
Qy 301 TYHNPADFITEVAGSEYGDLPNMLFRAVQNGLCMAEKKSSPEKNEVPAPCPPEVD 360
Db 301 TYHNPADFITEVAGSEYGDLPNMLFRAVQNGLCMAEKKSSPEKNEVPAPCPPEVD 360
Qy 361 IESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMSHVIVIGLILYHIGDDASKV 420
Db 361 IESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMSHVIVIGLILYHIGDDASKV 420
Qy 421 FNTGCLFFSNLFLMFAALMPTVLTPLFMAVMEHNLNWSLKAYLAKTWADVPFQV 480
Db 421 FNTGCLFFSNLFLMFAALMPTVLTPLFMAVMEHNLNWSLKAYLAKTWADVPFQV 480
Qy 481 VCPVYCSIVYWTGQPAETSRLFLFSALATATALVAQSLGLLIGASNSLQVATFVGPV 540
Db 481 VCPVYCSIVYWTGQPAETSRLFLFSALATATALVAQSLGLLIGASNSLQVATFVGPV 540
Qy 541 TAIPVLLFSGFFVSFKTIPTYLQSSYLSYVRVYGFEGVILTIYGMERGLDTCLEERC 600
Db 541 TAIPVLLFSGFFVSFKTIPTYLQSSYLSYVRVYGFEGVILTIYGMERGLDTCLEERC 600
Qy 601 EPOSILRALDVEDAKLYMDFLVLGIFPLALRLAYLVLRVRSKER 646
Db 601 EPOSILRALDVEDAKLYMDFLVLGIFPLALRLAYLVLRVRSKER 646

RESULT 2
ABB57112

ID ABB57112 standard; Protein: 666 AA.
XX
XX ABB57112;
XX
XX 07-MAR-2002 (first entry)
XX
XX Mouse ischaemic condition related protein sequence SEQ ID NO:255.
DE
XX
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.
XX
XX Mus musculus.
OS
XX
XX WO2001188188-A2.
PN
XX
XX 22-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001WO-JP04192.
PF
XX
XX 18-MAY-2000; 2000JP-0145977.
PR
XX
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
PA
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
PI
XX
XX WPI: 2002-034733/04.
DR
XX
XX N-PSDB: ABI99363.
DR
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
PT
XX
XX Claim 2; Page 719-722; 2690pp; English.
PS
XX
XX The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI9912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischaemic condition-improving
CC drugs or therapeutics for ischaemic diseases. ABI9913 and ABI9914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
XX Sequence 666 AA;
SQ
Query Match 70.38; Score 2353; DB 23; Length 666;
Best Local Similarity 72.28; Pred. No. 1.1e-218;
Matches 451; Conservative 80; Mismatches 90; Indels 4; Gaps 2;
Qy 26 DGAEPVPLVTHLKKVENHTEAQRSHLPKRSAVDIEFVELSYVREGPCWKRKGYKTL 85
Db 42 DEVETDLLNGHLKKVDNNFTAEQRFSLPRRAVNIEFKDLSYVPGCPWKKKGYKTL 101
Qy 86 KCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRESGMKGOILVNGRPRLTFRKMSY 145
Db 102 KGISGFNSGELVAIMGSPGAGKSTLMNLAGYRETGMKGAVALINGMPDLRCFRKVCY 161
Qy 146 IMODDMLLPHLTVLEAMVSNLNTENPDVKNDLVTEILTALGLMSCSHTTALLSGGQ 205
Db 162 IMODDMLLPHLTVOEAMVSAHLKQEGREVMKEILTALGLLPCANTRTGLSGGQ 221
Qy 206 RKRLATALELVNPPVPMFDEPTSGLDSASCFQVSLMKSLAQGGRTICTTHQPSAKLF 265
Db 222 RKRLATALELVNPPVPMFDEPTSGLDSASCFQVSLMKSLAQGGRTICTTHQPSAKLF 281
Qy 266 EMFDKLYLSQGCIFKGVVTVNLIPYLKGLGLHCPVHNPADEFIEVASGEYGDLPNMLF 325

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:14:34 ; Search time 73 Seconds
(without alignments)
1179.177 Million cell updates/sec

Title: US-10-072-621-9

Perfect score: 3347

Sequence: 1 MAEKALEAVCGGLPGAVAM.....FLALRLAYLVLRVVKSR 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

rched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Genesec_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3339	99.8	646	AA014186	Human transporter
2	2353	70.3	666	AB571112	Mouse ischaemic CO
3	2299	68.7	935	AAW78589	Human protein SEQ
4	2299	68.7	935	AAW00994	Human bone marrow
5	1723	51.5	832	AB559444	Drosophila melanog
6	1513	45.2	689	AAV78980	Silkworm Bm white
7	1322	39.5	689	AB561867	Drosophila melanog
8	1241.5	37.1	620	AB59648	Drosophila melanog
9	1063.5	34.8	812	AB564565	Drosophila melanog
10	1066.5	32.5	609	AB564566	Drosophila melanog

11	1067	31.9	615	22	AB564724	Drosophila melanog
12	1051	31.4	788	22	AB567949	Drosophila melanog
13	1016.5	30.4	643	22	AB560914	Drosophila melanog
14	1016.5	30.4	643	22	AB566977	Drosophila melanog
15	1016.5	30.4	643	22	AB566982	Drosophila melanog
16	888	26.5	597	22	AB563205	Drosophila melanog
17	879.5	26.3	687	22	AB559384	Drosophila melanog
18	845.5	25.3	687	21	AAV78981	Silkworm Bm white
19	767.5	22.9	625	21	AAV18080	Arabidopsis thalia
20	767.5	22.9	632	21	AAV18079	Arabidopsis thalia
21	767.5	22.9	648	21	AAV18078	Arabidopsis thalia
22	719	21.5	657	23	AB507272	Arabidopsis thalia
23	717	21.4	655	23	AB507273	Murine BCRP (hBCRP)
24	712	21.3	655	22	AA560104	Human BCRP (hBCRP)
25	712	21.3	655	23	AAU80028	Human transport pr
26	712	21.3	655	23	AA014781	Human ABCG2.. Homo
27	710	21.2	655	23	AAU80029	Human BCRP protein
28	710	21.2	663	20	AAV15221	Human ABCG2 mutant
29	708	21.2	665	23	AA014783	Breast Cancer Resi
30	707	21.1	665	23	AA014782	Human BCRP-related
31	704	21.0	655	21	AAV5365	ATP-binding casset
32	704	21.0	655	22	AAU04348	Human BCRP/MXR/ABC
33	704	21.0	655	23	AB507270	Human BCRP (hBCRP)
34	684	20.4	666	22	AB560524	Drosophila melanog
35	667.5	19.9	729	22	AB559797	Drosophila melanog
36	624.5	18.7	604	20	AAW73627	Human secreted pro
37	624.5	18.7	604	23	ABP61858	Human polypeptide
38	613	18.4	134	23	ABP51389	Human MDDT SEQ ID
39	613	18.3	649	21	AA043100	Arabidopsis thalia
40	613	18.3	653	21	AA043099	Arabidopsis thalia
41	586.5	17.5	651	23	AAU96989	Human ABCG5 mutant
42	586.5	17.5	651	23	AAU96990	Human ABCG5 mutant
43	586	17.5	616	21	AA050555	Arabidopsis thalia
44	585.5	17.5	651	23	AAU96984	Arabidopsis thali
45	585.5	17.5	651	23	AAE13290	Human sitosterolae

ALIGNMENTS

RESULT 1
AA014186
ID AA014186 standard; Protein: 646 AA.
XX AA014186;
AC AC
XX 03-MAY-2002 (first entry)
DT Human transporter and ion channel TRICH-3.
DE Human transporter and ion channel: TRICH; transport disorder;
KW neurological disorder; muscle disorder; immunological disorder;
KW cell proliferative disorder; neuroprotective; nontropic;
KW cerebroprotective; immunosuppressive; cytostatic; respiratory; muscular;
KW gene therapy.
XX Homo sapiens.
XX WO200204520-A2.
PN 17-JAN-2002.
PD 05-JUL-2001; 2001WO-US21448.
PF 07-JUL-2000; 2000US-216547P.
PR 14-JUL-2000; 2000US-218232P.
PR 21-JUL-2000; 2000US-220112P.
PR 28-JUL-2000; 2000US-221839P.
XX (INCY-) INCYTE GENOMICS INC.
PA Raumann BE, Thornton M, Ding L, Yue H, Tang YT, Harland L;
XX Burford N, Greene BD, Sanjanwala MS, Baughn MR, Yao MG, Yang J;
PI

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:36:29 ; Search time 14 Seconds
(without alignments)
1913.836 Million cell updates/sec

Title: US-10-072-621-9

Perfect score: 3347

Sequence: 1 MAEKALEAVGCGLGPGAVAM.....FLALRLLAYLVLYRVKRSER 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

rched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3309	98.9	646	1 ABG4_HUMAN	Q9h172 homo sapien
2	2356	70.4	678	1 ABG1_HUMAN	P45844 homo sapien
3	2353	70.3	666	1 ABG1_MOUSE	Q64343 mus musculus
4	879.5	26.3	687	1 WHIT_DROME	P10090 drosophila
5	806	24.1	679	1 WHIT_CERCA	Q17320 ceratitidis c
6	794.5	23.7	677	1 WHIT_LUCCU	Q05360 lucilia cup
7	779	23.3	695	1 WHIT_ANOGA	Q27256 anopheles g
8	736.5	22.0	709	1 WHIT_ANOAL	Q16928 anopheles a
9	723.5	21.6	598	1 YPC3_CAEEL	Q11180 caenorhabdi
10	712	21.3	655	1 ABG2_HUMAN	Q9unq0 homo sapien
11	667	19.9	666	1 SCRT_DROME	P45843 drosophila
12	595	17.8	1049	1 ADP1_YEAST	P23371 saccharomyc
13	585.5	17.5	651	1 ABG5_HUMAN	Q9h222 homo sapien
14	584	17.4	672	1 ABG8_RAT	P58428 rattus norv
15	577.5	17.3	652	1 ABG5_MOUSE	Q99pe8 mus musculus
16	577	17.2	610	1 YQ5C_CAEEL	Q09466 caenorhabdi
17	565	16.9	1294	1 YOH5_YEAST	Q08234 saccharomyc
18	563.5	16.8	652	1 ABG5_RAT	Q99pe7 rattus norv
19	563.5	16.8	673	1 ABG8_HUMAN	Q9h221 homo sapien
20	557.5	16.7	673	1 ABG8_MOUSE	Q9dbm0 mus musculus
21	531.5	15.9	650	1 ABG3_MOUSE	Q99pe81 mus musculus
22	483	14.4	1333	1 YN99_YEAST	P53756 saccharomyc
23	472	14.1	1529	1 PDRF_YEAST	Q04182 saccharomyc
24	469	14.0	1530	1 BFP1_SCHPO	P41820 schizosacch
25	466.5	13.9	1511	1 PDRC_YEAST	Q02785 saccharomyc
26	462.5	13.8	1564	1 PDRA_YEAST	P51533 saccharomyc
27	459	13.7	1501	1 SNQ2_YEAST	P32568 saccharomyc
28	455	13.6	675	1 BROW_DROME	P12428 drosophila
29	445.5	13.3	1511	1 PDR5_YEAST	P33302 saccharomyc
30	437.5	13.1	1501	1 CDR3_CANAL	Q04690 candida alb
31	429	12.8	1490	1 CDR4_CANAL	O74676 candida alb
32	424.5	12.7	1499	1 CDR2_CANAL	P78595 candida alb
33	413.5	12.4	1501	1 CDR1_CANAL	P43071 candida alb

RESULT 1

ID	ABG4_HUMAN	STANDARD	PRT	646 AA
AC	Q9H172			
DT	15-JUN-2002	(Rel. 41, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	ATP-binding cassette, sub-family G, member 4.			
GN	ABCG4 OR WHITE2			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid:9606			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21518231; PubMed=11606068;			
RA	Engel T., Lorkowski S., Lueken A., Rust S., Schlueter B., Berger G.,			
RA	Cullen P., Assmann G.;			
RT	"The human ABCG4 gene is regulated by oxysterols and retinoids in			
RT	monocyte-derived macrophages."			
RL	Biochem. Biophys. Res. Commun. 288:483-488(2001).			
RN	[2]			
RP	SEQUENCE OF 20-646 FROM N.A.			
RC	TISSUE=Dorsal root ganglion;			
RA	Oldfield S., Lowry C.A., Lightman S.L.;			
RT	"Cloning and expression of a novel mammalian white family			
RT	ABC transporter: WHITE2."			
CC	Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.			
CC	!- FUNCTION: May be involved in macrophage lipid homeostasis.			
CC	!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).			
CC	!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AJ300465; CAC17140.1;			
DR	Genew; HGNC:13884; ABCG4.			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR003439; ABC_transportr.			
DR	Pfam; PF00005; ABC_tran; 2.			
DR	ProDom; PD000006; ABC_transportr; 1.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS00211; ABC_TRANSPORTER; 1.			
KW	ATP-binding; Glycoprotein; Transmembrane; Transport.			
FT	DOMAIN 1 393			
FT	TRANSMEM 394 414			
FT	DOMAIN 415 425			
FT	TRANSMEM 426 446			
FT	DOMAIN 447 472			
FT	TRANSMEM 473 493			

ALIGNMENTS

Q24739 drosophila
P40550 saccharomyc
P50332 rhizobium g
Q58903 methanococ
P08720 rhizobium l
P54537 bacillus su
Q05067 rhizobium s
Q05067 anabaena sp
P10346 escherichia
P75612 mycoplasma
Q9wy17 thermotoga
P44986 haemophilus

34 411.5 12.3 668 1 BROW_DROVI
35 405 12.1 1410 1 PDRB_YEAST
36 263.5 7.9 347 1 NOD1_RHIGA
37 258.5 7.7 224 1 YF08_METJA
38 254 7.6 311 1 NOD1_RHILV
39 253.5 7.6 240 1 YQIZ_BACSU
40 249 7.4 371 1 Y4OS_RHISN
41 244 7.3 335 1 Y7I9_ANASP
42 243.5 7.3 240 1 GLNQ_ECOLI
43 242 7.2 465 1 Y065_MYCPN
44 239 7.1 234 1 Y352_THEMA
45 238.5 7.1 229 1 THIQ_HAEIN

```

FT DOMAIN 494 503 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 504 524 4 (POTENTIAL).
FT DOMAIN 525 532 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 533 553 5 (POTENTIAL).
FT DOMAIN 554 617 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 618 638 6 (POTENTIAL).
FT DOMAIN 639 646 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 102 109 ATP (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLCNAC..) (POTENTIAL).
SQ SEQUENCE 646 AA; 71895 MW; 9CCEC6E150772611 CRC64;

Query Match 98.9%; Score 3309; DB 1; Length 646;
Best Local Similarity 98.9%; Pred. No. 3;le-214;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVNHTEAQRSHLPKRSVD 60
Db 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVNHTEAQRSHLPKRSVD 60
QY 61 IEFVELSVREGPCWRKRGYKTLKLSGKFCRRELIGIMGPSGAGKSTFMNLAGYRE 120
61 IEFVELSVREGPCWRKRGYKTLKLSGKFCRRELIGIMGPSGAGKSTFMNLAGYRE 120
121 SGMKQOILVNGPRELFRKMSYIMQDDMLLPHLTVLEAMWVANLNTENPDVKNDL 180
121 SGMKQOILVNGPRELFRKMSYIMQDDMLLPHLTVLEAMWVANLNTENPDVKNDL 180
QY 181 VTEILTALGLMSCSTRTALLSGGQRKLATALELVNPNPVFFDEPTSGDLSASCFOV 240
181 VTEILTALGLMSCSTRTALLSGGQRKLATALELVNPNPVFFDEPTSGDLSASCFOV 240
QY 241 SLMSLAQGGRTIICTIHOPSAKLFEMEDKLYLSQGGCIPKGVVNTNLPYKLGILHCP 300
241 SLMSLAQGGRTIICTIHOPSAKLFEMEDKLYLSQGGCIPKGVVNTNLPYKLGILHCP 300
QY 301 TYHNPADRIIEVASGEYDGLNPLFRAVONGLCMAEKSSPEKNEVPAPCPPEVD 360
301 TYHNPADRIIEVASGEYDGLNPLFRAVONGLCMAEKSSPEKNEVPAPCPPEVD 360
QY 361 IESHFTFATSTLQFLEKFRFLSLRDTLTLHFMHSHVIGVLGLLHIGDDASKV 420
361 IESHFTFATSTLQFLEKFRFLSLRDTLTLHFMHSHVIGVLGLLHIGDDASKV 420
QY 421 FNNTCGLFFSMLFLMFAALMPTVLTFLPFAVMEHLYNYSKAYYLAKTMADVPFQV 480
421 FNNTCGLFFSMLFLMFAALMPTVLTFLPFAVMEHLYNYSKAYYLAKTMADVPFQV 480
QY 481 VCPVVCYSIVVMTGQPAETSRFLFSALATATATVAOSGLLIGAAANSLOVATFVGPV 540
481 VCPVVCYSIVVMTGQPAETSRFLFSALATATATVAOSGLLIGAAANSLOVATFVGPV 540
QY 541 TAIPVLLFSGFFVSFKTIPTIQLWSSYLSYVRYGEGVILTYGMERGLDLCLEERCPCR 600
541 TAIPVLLFSGFFVSFKTIPTIQLWSSYLSYVRYGEGVILTYGMERGLDLCLEERCPCR 600
QY 601 EPQSTILRALDVEDAKLYMDLVLGIFFLALRLAYLVLYRYVKSR 646
601 EPQSTILRALDVEDAKLYMDLVLGIFFLALRLAYLVLYRYVKSR 646

```

RESULT 2

```

ABG1_HUMAN STANDARD; PRT; 678 AA.
ID ABG1_HUMAN Q9BXL0; Q9BXL1; Q9BXL2; Q9BXL3; Q9BXL4; Q9BXL6; Q9BXL7;
AC P45844; Q9BXL0; Q9BXL1; Q9BXL2; Q9BXL3; Q9BXL4; Q9BXL6; Q9BXL7;
AC Q9BXL8; Q9BXL9;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
DE (ATP-binding cassette transporter 8).
GN ABCG1 OR ABC8 OR WHR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
RC TISSUE=Retina;
RX MEDLINE=96256850; PubMed=8659545;
RA Chen H.M., Rossier C., Lalliot M.D., Lynn A., Chakravarti A.,
Perrin G., Antonarakis S.E.;
RT "Cloning of the cDNA for a human homologue of the Drosophila white
gene and mapping to chromosome 21q22.3.";
RL Am. J. Hum. Genet. 59:66-75(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagan E.,
Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20408883; PubMed=10950923;
RA Berry A., Scott H.S., Kudoh J., Tallor I., Korostishevsky M.,
Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,
Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
Antonarakis S.E., Bonne-Tamir B.;
RT "Refined localization of autosomal recessive nonsyndromic deafness
DFNB10 locus using 34 novel microsatellite markers, genomic
structure, and exclusion of six known genes in the region.";
RL Genomics 68:22-29(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21192304; PubMed=11279031;
RA Porsch-Oecueruemeiz M., Langmann T., Heimerl S., Borsukova H.,
Kaminski W.E., Drobniak W., Honer C., Schumacher C., Schmitz G.;
RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
expression and a modulator of cellular lipid efflux.";
RL J. Biol. Chem. 276:12427-12433(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).
RX MEDLINE=21092576; PubMed=11162488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
Assmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [6]
RP SEQUENCE OF 33-678 FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97186700; PubMed=9034316;
RA Croop J.M., Filler G.E., Fletcher J.A., Lux M.L., Raab E.,
Goldenson D., Arciniegas S., Son D., Wu R.;
RT "Isolation and characterization of a mammalian homologue of the
Drosophila white gene.";
RL Gene 185:77-85(1997).
RN [7]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE=20261604; PubMed=10799558;
RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
lipid-loaded macrophages. A transcriptional role for specific
oxysterols.";
RL J. Biol. Chem. 275:14700-14707(2000).

```

RN INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=2010556; PubMed=10639163;
 RA Klucken J., Buechler C., Orso E., Kaminski W.E.,
 RA Porsch-Oezcuernem M., Liebisch G., Kapinsky M., Diederich W.,
 RA Drobnik W., Dean M., Altknecht R., Schmitz G.;
 RT "ABCG1 (ABCG8), the human homolog of the brospophila white gene, is a
 RT regulator of macrophage cholesterol and phospholipid transport.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
 RN [9]
 RP REVIEW
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 CC an active component of the macrophage lipid export complex. Could
 CC also be involved in intracellular lipid transport processes. The
 CC role in cellular lipid homeostasis may not be limited to
 CC macrophages.
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners
 CC of the ABCG subfamily.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC Predominantly localized in the intracellular compartments mainly
 CC associated with the endoplasmic reticulum (ER) and Golgi
 CC membranes.
 CC -!- ALTERNATIVE PRODUCTS: At least 7 isoforms; 1 (shown here), 2/J,
 CC 3/ABDE, 4/G, 5/F, 6/HI and 7/C; are produced by alternative
 CC splicing and alternative initiation.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
 CC -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
 CC cholesterol influx. Conversely, mRNA and protein expression are
 CC suppressed by lipid efflux. Induction is mediated by the
 CC liver x receptor/retinoid x receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X91249; CAA62631.1; ALT_INIT.
 DR EMBL; AP001746; BAA95530.1; ALT_INIT.
 DR EMBL; AB038161; BAB13728.2; ALT_INIT.
 DR EMBL; AJ289137; CAC00730.1; JOINED.
 DR EMBL; AJ289138; CAC00730.1; JOINED.
 DR EMBL; AJ289139; CAC00730.1; JOINED.
 DR EMBL; AJ289140; CAC00730.1; JOINED.
 DR EMBL; AJ289141; CAC00730.1; JOINED.
 DR EMBL; AJ289142; CAC00730.1; JOINED.
 DR EMBL; AJ289143; CAC00730.1; JOINED.
 DR EMBL; AJ289144; CAC00730.1; JOINED.
 DR EMBL; AJ289145; CAC00730.1; JOINED.
 DR EMBL; AJ289146; CAC00730.1; JOINED.
 DR EMBL; AJ289147; CAC00730.1; JOINED.
 DR EMBL; AJ289148; CAC00730.1; JOINED.
 DR EMBL; AJ289149; CAC00730.1; JOINED.
 DR EMBL; AJ289150; CAC00730.1; JOINED.
 DR EMBL; AJ289151; CAC00730.1; JOINED.
 DR EMBL; AF323658; AAK28836.1; JOINED.
 DR EMBL; AF323644; AAK28836.1; JOINED.
 DR EMBL; AF323645; AAK28836.1; JOINED.
 DR EMBL; AF323646; AAK28836.1; JOINED.
 DR EMBL; AF323647; AAK28836.1; JOINED.
 DR EMBL; AF323648; AAK28836.1; JOINED.
 DR EMBL; AF323649; AAK28836.1; JOINED.
 DR EMBL; AF323650; AAK28836.1; JOINED.
 DR EMBL; AF323651; AAK28836.1; JOINED.
 DR EMBL; AF323652; AAK28836.1; JOINED.
 DR EMBL; AF323653; AAK28836.1; JOINED.
 DR EMBL; AF323654; AAK28836.1; JOINED.
 DR EMBL; AF323655; AAK28836.1; JOINED.
 DR EMBL; AF323656; AAK28836.1; JOINED.
 DR EMBL; AF323657; AAK28836.1; JOINED.
 DR EMBL; AF323658; AAK28836.1; JOINED.
 DR EMBL; AF323659; AAK28836.1; JOINED.
 DR EMBL; AF323660; AAK28836.1; JOINED.
 DR EMBL; AF323661; AAK28836.1; JOINED.
 DR EMBL; AF323662; AAK28836.1; JOINED.
 DR EMBL; AF323663; AAK28836.1; JOINED.
 DR EMBL; AF323664; AAK28836.1; JOINED.
 DR EMBL; AF323665; AAK28836.1; JOINED.
 DR EMBL; AF323666; AAK28836.1; JOINED.
 DR EMBL; AF323667; AAK28836.1; JOINED.
 DR EMBL; AF323668; AAK28836.1; JOINED.
 DR EMBL; AF323669; AAK28836.1; JOINED.
 DR EMBL; AF323670; AAK28836.1; JOINED.
 DR EMBL; AF323671; AAK28836.1; JOINED.
 DR EMBL; AF323672; AAK28836.1; JOINED.
 DR EMBL; AF323673; AAK28836.1; JOINED.
 DR EMBL; AF323674; AAK28836.1; JOINED.
 DR EMBL; AF323675; AAK28836.1; JOINED.
 DR EMBL; AF323676; AAK28836.1; JOINED.
 DR EMBL; AF323677; AAK28836.1; JOINED.
 DR EMBL; AF323678; AAK28836.1; JOINED.
 DR EMBL; AF323679; AAK28836.1; JOINED.
 DR EMBL; AF323680; AAK28836.1; JOINED.
 DR EMBL; AF323681; AAK28836.1; JOINED.
 DR EMBL; AF323682; AAK28836.1; JOINED.
 DR EMBL; AF323683; AAK28836.1; JOINED.
 DR EMBL; AF323684; AAK28836.1; JOINED.
 DR EMBL; AF323685; AAK28836.1; JOINED.
 DR EMBL; AF323686; AAK28836.1; JOINED.
 DR EMBL; AF323687; AAK28836.1; JOINED.
 DR EMBL; AF323688; AAK28836.1; JOINED.
 DR EMBL; AF323689; AAK28836.1; JOINED.
 DR EMBL; AF323690; AAK28836.1; JOINED.
 DR EMBL; AF323691; AAK28836.1; JOINED.
 DR EMBL; AF323692; AAK28836.1; JOINED.
 DR EMBL; AF323693; AAK28836.1; JOINED.
 DR EMBL; AF323694; AAK28836.1; JOINED.
 DR EMBL; AF323695; AAK28836.1; JOINED.
 DR EMBL; AF323696; AAK28836.1; JOINED.
 DR EMBL; AF323697; AAK28836.1; JOINED.
 DR EMBL; AF323698; AAK28836.1; JOINED.
 DR EMBL; AF323699; AAK28836.1; JOINED.
 DR EMBL; AF323700; AAK28836.1; JOINED.
 DR EMBL; AF323701; AAK28836.1; JOINED.
 DR EMBL; AF323702; AAK28836.1; JOINED.
 DR EMBL; AF323703; AAK28836.1; JOINED.
 DR EMBL; AF323704; AAK28836.1; JOINED.
 DR EMBL; AF323705; AAK28836.1; JOINED.
 DR EMBL; AF323706; AAK28836.1; JOINED.
 DR EMBL; AF323707; AAK28836.1; JOINED.
 DR EMBL; AF323708; AAK28836.1; JOINED.
 DR EMBL; AF323709; AAK28836.1; JOINED.
 DR EMBL; AF323710; AAK28836.1; JOINED.
 DR EMBL; AF323711; AAK28836.1; JOINED.
 DR EMBL; AF323712; AAK28836.1; JOINED.
 DR EMBL; AF323713; AAK28836.1; JOINED.
 DR EMBL; AF323714; AAK28836.1; JOINED.
 DR EMBL; AF323715; AAK28836.1; JOINED.
 DR EMBL; AF323716; AAK28836.1; JOINED.
 DR EMBL; AF323717; AAK28836.1; JOINED.
 DR EMBL; AF323718; AAK28836.1; JOINED.
 DR EMBL; AF323719; AAK28836.1; JOINED.
 DR EMBL; AF323720; AAK28836.1; JOINED.
 DR EMBL; AF323721; AAK28836.1; JOINED.
 DR EMBL; AF323722; AAK28836.1; JOINED.
 DR EMBL; AF323723; AAK28836.1; JOINED.
 DR EMBL; AF323724; AAK28836.1; JOINED.
 DR EMBL; AF323725; AAK28836.1; JOINED.
 DR EMBL; AF323726; AAK28836.1; JOINED.
 DR EMBL; AF323727; AAK28836.1; JOINED.
 DR EMBL; AF323728; AAK28836.1; JOINED.
 DR EMBL; AF323729; AAK28836.1; JOINED.
 DR EMBL; AF323730; AAK28836.1; JOINED.
 DR EMBL; AF323731; AAK28836.1; JOINED.
 DR EMBL; AF323732; AAK28836.1; JOINED.
 DR EMBL; AF323733; AAK28836.1; JOINED.
 DR EMBL; AF323734; AAK28836.1; JOINED.
 DR EMBL; AF323735; AAK28836.1; JOINED.
 DR EMBL; AF323736; AAK28836.1; JOINED.
 DR EMBL; AF323737; AAK28836.1; JOINED.
 DR EMBL; AF323738; AAK28836.1; JOINED.
 DR EMBL; AF323739; AAK28836.1; JOINED.
 DR EMBL; AF323740; AAK28836.1; JOINED.
 DR EMBL; AF323741; AAK28836.1; JOINED.
 DR EMBL; AF323742; AAK28836.1; JOINED.
 DR EMBL; AF323743; AAK28836.1; JOINED.
 DR EMBL; AF323744; AAK28836.1; JOINED.
 DR EMBL; AF323745; AAK28836.1; JOINED.
 DR EMBL; AF323746; AAK28836.1; JOINED.
 DR EMBL; AF323747; AAK28836.1; JOINED.
 DR EMBL; AF323748; AAK28836.1; JOINED.
 DR EMBL; AF323749; AAK28836.1; JOINED.
 DR EMBL; AF323750; AAK28836.1; JOINED.
 DR EMBL; AF323751; AAK28836.1; JOINED.
 DR EMBL; AF323752; AAK28836.1; JOINED.
 DR EMBL; AF323753; AAK28836.1; JOINED.
 DR EMBL; AF323754; AAK28836.1; JOINED.
 DR EMBL; AF323755; AAK28836.1; JOINED.
 DR EMBL; AF323756; AAK28836.1; JOINED.
 DR EMBL; AF323757; AAK28836.1; JOINED.
 DR EMBL; AF323758; AAK28836.1; JOINED.
 DR EMBL; AF323759; AAK28836.1; JOINED.
 DR EMBL; AF323760; AAK28836.1; JOINED.
 DR EMBL; AF323761; AAK28836.1; JOINED.
 DR EMBL; AF323762; AAK28836.1; JOINED.
 DR EMBL; AF323763; AAK28836.1; JOINED.
 DR EMBL; AF323764; AAK28836.1; JOINED.
 DR EMBL; AF323765; AAK28836.1; JOINED.
 DR EMBL; AF323766; AAK28836.1; JOINED.
 DR EMBL; AF323767; AAK28836.1; JOINED.
 DR EMBL; AF323768; AAK28836.1; JOINED.
 DR EMBL; AF323769; AAK28836.1; JOINED.
 DR EMBL; AF323770; AAK28836.1; JOINED.
 DR EMBL; AF323771; AAK28836.1; JOINED.
 DR EMBL; AF323772; AAK28836.1; JOINED.
 DR EMBL; AF323773; AAK28836.1; JOINED.
 DR EMBL; AF323774; AAK28836.1; JOINED.
 DR EMBL; AF323775; AAK28836.1; JOINED.
 DR EMBL; AF323776; AAK28836.1; JOINED.
 DR EMBL; AF323777; AAK28836.1; JOINED.
 DR EMBL; AF323778; AAK28836.1; JOINED.
 DR EMBL; AF323779; AAK28836.1; JOINED.
 DR EMBL; AF323780; AAK28836.1; JOINED.
 DR EMBL; AF323781; AAK28836.1; JOINED.
 DR EMBL; AF323782; AAK28836.1; JOINED.
 DR EMBL; AF323783; AAK28836.1; JOINED.
 DR EMBL; AF323784; AAK28836.1; JOINED.
 DR EMBL; AF323785; AAK28836.1; JOINED.
 DR EMBL; AF323786; AAK28836.1; JOINED.
 DR EMBL; AF323787; AAK28836.1; JOINED.
 DR EMBL; AF323788; AAK28836.1; JOINED.
 DR EMBL; AF323789; AAK28836.1; JOINED.
 DR EMBL; AF323790; AAK28836.1; JOINED.
 DR EMBL; AF323791; AAK28836.1; JOINED.
 DR EMBL; AF323792; AAK28836.1; JOINED.
 DR EMBL; AF323793; AAK28836.1; JOINED.
 DR EMBL; AF323794; AAK28836.1; JOINED.
 DR EMBL; AF323795; AAK28836.1; JOINED.
 DR EMBL; AF323796; AAK28836.1; JOINED.
 DR EMBL; AF323797; AAK28836.1; JOINED.
 DR EMBL; AF323798; AAK28836.1; JOINED.
 DR EMBL; AF323799; AAK28836.1; JOINED.
 DR EMBL; AF323800; AAK28836.1; JOINED.
 DR EMBL; AF323801; AAK28836.1; JOINED.
 DR EMBL; AF323802; AAK28836.1; JOINED.
 DR EMBL; AF323803; AAK28836.1; JOINED.
 DR EMBL; AF323804; AAK28836.1; JOINED.
 DR EMBL; AF323805; AAK28836.1; JOINED.
 DR EMBL; AF323806; AAK28836.1; JOINED.
 DR EMBL; AF323807; AAK28836.1; JOINED.
 DR EMBL; AF323808; AAK28836.1; JOINED.
 DR EMBL; AF323809; AAK28836.1; JOINED.
 DR EMBL; AF323810; AAK28836.1; JOINED.
 DR EMBL; AF323811; AAK28836.1; JOINED.
 DR EMBL; AF323812; AAK28836.1; JOINED.
 DR EMBL; AF323813; AAK28836.1; JOINED.
 DR EMBL; AF323814; AAK28836.1; JOINED.
 DR EMBL; AF323815; AAK28836.1; JOINED.
 DR EMBL; AF323816; AAK28836.1; JOINED.
 DR EMBL; AF323817; AAK28836.1; JOINED.
 DR EMBL; AF323818; AAK28836.1; JOINED.
 DR EMBL; AF323819; AAK28836.1; JOINED.
 DR EMBL; AF323820; AAK28836.1; JOINED.
 DR EMBL; AF323821; AAK28836.1; JOINED.
 DR EMBL; AF323822; AAK28836.1; JOINED.
 DR EMBL; AF323823; AAK28836.1; JOINED.
 DR EMBL; AF323824; AAK28836.1; JOINED.
 DR EMBL; AF323825; AAK28836.1; JOINED.
 DR EMBL; AF323826; AAK28836.1; JOINED.
 DR EMBL; AF323827; AAK28836.1; JOINED.
 DR EMBL; AF323828; AAK28836.1; JOINED.
 DR EMBL; AF323829; AAK28836.1; JOINED.
 DR EMBL; AF323830; AAK28836.1; JOINED.
 DR EMBL; AF323831; AAK28836.1; JOINED.
 DR EMBL; AF323832; AAK28836.1; JOINED.
 DR EMBL; AF323833; AAK28836.1; JOINED.
 DR EMBL; AF323834; AAK28836.1; JOINED.
 DR EMBL; AF323835; AAK28836.1; JOINED.
 DR EMBL; AF323836; AAK28836.1; JOINED.
 DR EMBL; AF323837; AAK28836.1; JOINED.
 DR EMBL; AF323838; AAK28836.1; JOINED.
 DR EMBL; AF323839; AAK28836.1; JOINED.
 DR EMBL; AF323840; AAK28836.1; JOINED.
 DR EMBL; AF323841; AAK28836.1; JOINED.
 DR EMBL; AF323842; AAK28836.1; JOINED.
 DR EMBL; AF323843; AAK28836.1; JOINED.
 DR EMBL; AF323844; AAK28836.1; JOINED.
 DR EMBL; AF323845; AAK28836.1; JOINED.
 DR EMBL; AF323846; AAK28836.1; JOINED.
 DR EMBL; AF323847; AAK28836.1; JOINED.
 DR EMBL; AF323848; AAK28836.1; JOINED.
 DR EMBL; AF323849; AAK28836.1; JOINED.
 DR EMBL; AF323850; AAK28836.1; JOINED.
 DR EMBL; AF323851; AAK28836.1; JOINED.
 DR EMBL; AF323852; AAK28836.1; JOINED.
 DR EMBL; AF323853; AAK28836.1; JOINED.
 DR EMBL; AF323854; AAK28836.1; JOINED.
 DR EMBL; AF323855; AAK28836.1; JOINED.
 DR EMBL; AF323856; AAK28836.1; JOINED.
 DR EMBL; AF323857; AAK28836.1; JOINED.
 DR EMBL; AF323858; AAK28836.1; JOINED.
 DR EMBL; AF323859; AAK28836.1; JOINED.
 DR EMBL; AF323860; AAK28836.1; JOINED.
 DR EMBL; AF323861; AAK28836.1; JOINED.
 DR EMBL; AF323862; AAK28836.1; JOINED.
 DR EMBL; AF323863; AAK28836.1; JOINED.
 DR EMBL; AF323864; AAK28836.1; JOINED.
 DR EMBL; AF323865; AAK28836.1; JOINED.
 DR EMBL; AF323866; AAK28836.1; JOINED.
 DR EMBL; AF323867; AAK28836.1; JOINED.
 DR EMBL; AF323868; AAK28836.1; JOINED.
 DR EMBL; AF323869; AAK28836.1; JOINED.
 DR EMBL; AF323870; AAK28836.1; JOINED.
 DR EMBL; AF323871; AAK28836.1; JOINED.
 DR EMBL; AF323872; AAK28836.1; JOINED.
 DR EMBL; AF323873; AAK28836.1; JOINED.
 DR EMBL; AF323874; AAK28836.1; JOINED.
 DR EMBL; AF323875; AAK28836.1; JOINED.
 DR EMBL; AF323876; AAK28836.1; JOINED.
 DR EMBL; AF323877; AAK28836.1; JOINED.
 DR EMBL; AF323878; AAK28836.1; JOINED.
 DR EMBL; AF323879; AAK28836.1; JOINED.
 DR EMBL; AF323880; AAK28836.1; JOINED.
 DR EMBL; AF323881; AAK28836.1; JOINED.
 DR EMBL; AF323882; AAK28836.1; JOINED.
 DR EMBL; AF323883; AAK28836.1; JOINED.
 DR EMBL; AF323884; AAK28836.1; JOINED.
 DR EMBL; AF323885; AAK28836.1; JOINED.
 DR EMBL; AF323886; AAK28836.1; JOINED.
 DR EMBL; AF323887; AAK28836.1; JOINED.
 DR EMBL; AF323888; AAK28836.1; JOINED.
 DR EMBL; AF323889; AAK28836.1; JOINED.
 DR EMBL; AF323890; AAK28836.1; JOINED.
 DR EMBL; AF323891; AAK28836.1; JOINED.
 DR EMBL; AF323892; AAK28836.1; JOINED.
 DR EMBL; AF323893; AAK28836.1; JOINED.
 DR EMBL; AF323894; AAK28836.1; JOINED.
 DR EMBL; AF323895; AAK28836.1; JOINED.
 DR EMBL; AF323896; AAK28836.1; JOINED.
 DR EMBL; AF323897; AAK28836.1; JOINED.
 DR EMBL; AF323898; AAK28836.1; JOINED.
 DR EMBL; AF323899; AAK28836.1; JOINED.
 DR EMBL; AF323900; AAK28836.1; JOINED.
 DR EMBL; AF323901; AAK28836.1; JOINED.
 DR EMBL; AF323902; AAK28836.1; JOINED.
 DR EMBL; AF323903; AAK28836.1; JOINED.
 DR EMBL; AF323904; AAK28836.1; JOINED.
 DR EMBL; AF323905; AAK28836.1; JOINED.
 DR EMBL; AF323906; AAK28836.1; JOINED.
 DR EMBL; AF323907; AAK28836.1; JOINED.
 DR EMBL; AF323908; AAK28836.1; JOINED.
 DR EMBL; AF323909; AAK28836.1; JOINED.
 DR EMBL; AF323910; AAK28836.1; JOINED.
 DR EMBL; AF323911; AAK28836.1; JOINED.
 DR EMBL; AF323912; AAK28836.1; JOINED.
 DR EMBL; AF323913; AAK28836.1; JOINED.
 DR EMBL; AF323914; AAK28836.1; JOINED.
 DR EMBL; AF323915; AAK28836.1; JOINED.
 DR EMBL; AF323916; AAK28836.1; JOINED.
 DR EMBL; AF323917; AAK28836.1; JOINED.
 DR EMBL; AF323918; AAK28836.1; JOINED.
 DR EMBL; AF323919; AAK28836.1; JOINED.
 DR EMBL; AF323920; AAK28836.1; JOINED.
 DR EMBL; AF323921; AAK28836.1; JOINED.
 DR EMBL; AF323922; AAK28836.1; JOINED.
 DR EMBL; AF323923; AAK28836.1; JOINED.
 DR EMBL; AF323924; AAK28836.1; JOINED.
 DR EMBL; AF323925; AAK28836.1; JOINED.
 DR EMBL; AF323926; AAK28836.1; JOINED.
 DR EMBL; AF323927; AAK28836.1; JOINED.
 DR EMBL; AF323928; AAK28836.1; JOINED.
 DR EMBL; AF323929; AAK28836.1; JOINED.
 DR EMBL; AF323930; AAK28836.1; JOINED.
 DR EMBL; AF323931; AAK28836.1; JOINED.
 DR EMBL; AF323932; AAK28836.1; JOINED.
 DR EMBL; AF323933; AAK28836.1; JOINED.
 DR EMBL; AF323934; AAK28836.1; JOINED.
 DR EMBL; AF323935; AAK28836.1; JOINED.
 DR EMBL; AF323936; AAK28836.1; JOINED.
 DR EMBL; AF323937; AAK28836.1; JOINED.
 DR EMBL; AF323938; AAK28836.1; JOINED.
 DR EMBL; AF323939; AAK28836.1; JOINED.
 DR EMBL; AF323940; AAK28836.1; JOINED.
 DR EMBL; AF323941; AAK28836.1; JOINED.
 DR EMBL; AF323942; AAK28836.1; JOINED.
 DR EMBL; AF323943; AAK28836.1; JOINED.
 DR EMBL; AF323944; AAK28836.1; JOINED.
 DR EMBL; AF323945; AAK28836.1; JOINED.
 DR EMBL; AF323946; AAK28836.1; JOINED.
 DR EMBL; AF323947; AAK28836.1; JOINED.
 DR EMBL; AF323948; AAK28836.1; JOINED.
 DR EMBL; AF323949; AAK28836.1; JOINED.
 DR EMBL; AF323950; AAK28836.1; JOINED.
 DR EMBL; AF323951; AAK28836.1; JOINED.
 DR EMBL; AF323952; AAK28836.1; JOINED.
 DR EMBL; AF323953; AAK28836.1; JOINED.
 DR EMBL; AF323954; AAK28836.1; JOINED.
 DR EMBL; AF323955; AAK28836.1; JOINED.
 DR EMBL; AF323956; AAK28836.1; JOINED.
 DR EMBL; AF323957; AAK28836.1; JOINED.
 DR EMBL; AF323958; AAK28836.1; JOINED.
 DR EMBL; AF323959; AAK28836.1; JOINED.
 DR EMBL; AF323960; AAK28836.1; JOINED.
 DR EMBL; AF323961; AAK28836.1; JOINED.
 DR EMBL; AF323962; AAK28836.1; JOINED.
 DR EMBL; AF323963; AAK28836.1; JOINED.
 DR EMBL; AF323964; AAK28836.1; JOINED.
 DR EMBL; AF323965; AAK28836.1; JOINED.
 DR EMBL; AF323966; AAK28836.1; JOINED.
 DR EMBL; AF323967; AAK28836.1; JOINED.
 DR EMBL; AF323968; AAK28836.1; JOINED.
 DR EMBL; AF323969; AAK28836.1; JOINED.
 DR EMBL; AF323970; AAK28836.1; JOINED.
 DR EMBL; AF323971; AAK28836.1; JOINED.
 DR EMBL; AF323972; AAK28836.1; JOINED.
 DR EMBL; AF323973; AAK28836.1; JOINED.
 DR EMBL; AF323974; AAK28836.1; JOINED.
 DR EMBL; AF323975; AAK28836.1; JOINED.
 DR EMBL; AF323976; AAK28836.1; JOINED.
 DR EMBL; AF323977; AAK28836.1; JOINED.
 DR EMBL; AF323978; AAK28836.1; JOINED.
 DR EMBL; AF323979; AAK28836.1; JOINED.
 DR EMBL; AF323980; AAK28836.1; JOINED.
 DR EMBL; AF323981; AAK28836.1; JOINED.
 DR EMBL; AF323982; AAK28836.1; JOINED.
 DR EMBL; AF323983; AAK28836.1; JOINED.
 DR EMBL; AF323984; AAK28836.1; JOINED.
 DR EMBL; AF323985; AAK28836.1; JOINED.
 DR EMBL; AF323986; AAK28836.1; JOINED.
 DR EMBL; AF323987; AAK28836.1; JOINED.
 DR EMBL; AF323988; AAK28836.1; JOINED.
 DR EMBL; AF323989; AAK28836.1; JOINED.
 DR EMBL; AF323990; AAK28836.1; JOINED.
 DR EMBL; AF323991; AAK28836.1; JOINED.
 DR EMBL; AF323992; AAK28836.1; JOINED.
 DR EMBL; AF323993; AAK28836.1; JOINED.
 DR EMBL; AF323994; AAK28836.1; JOINED.
 DR EMBL; AF323995; AAK28836.1; JOINED.
 DR EMBL; AF323996; AAK28836.1; JOINED.
 DR EMBL; AF323997; AAK28836.1; JOINED.
 DR EMBL; AF323998; AAK28836.1; JOINED.
 DR EMBL; AF323999; AAK28836.1; JOINED.
 DR EMBL; AF324000; AAK28836.1; JOINED.
 DR EMBL; AF324001; AAK28836.1; JOINED.
 DR EMBL; AF324002; AAK28836.1; JOINED.
 DR

Db 282 ELFDQLVVLSSGQCVRGKVNLPVRLDGLNCLPHTYHNPAFVMEVASGEYGDQNSRLV 341
 QY 326 RAYQNGICMAEKSSPEKNEVPAPCPPEVDPIE---SHFATSTLTFQFCLKRTF 382
 Db 342 RAYREGCDADYXRDLGGDTDVNPFLLHWRPAEDSASMECHSFSASCLTFQFCLKRTF 401
 QY 383 LSLRDTVTLLRFMSHVIGVLGLYLLHIGDDASKVFNNTCGLFMSLFLMFAALMPT 442
 Db 402 LSLMRDSVTLHRLTSHIGIGLGLYLLIGNEAKKVLNSGFLFMSLFLMFAALMPT 461
 QY 443 VLFPLEMAYMEHLEHLYNYSKAYLAKTMADVPQVCPVYVCSIVYVMTQPAETSR 502
 Db 462 VLFPLEMSVFLREHLNLYSLKAYLAKTMADVPFOIMPFPVYVYVMTQPSDAVR 521
 QY 503 FLFLSALATATLVAQSLGLLGAASNSQVATFVGPVTAIPVLLFSGFVSEKTIPTYL 562
 Db 522 FVLFAALGTMTSLVAQSLGLLGAASNSQVATFVGPVTAIPVLLFSGFVSEKTIPTYL 581
 QY 563 QNSYSLSYVRGEGVILTYIGMERGLTC-LSERCPPFPQSLRALDVEDAKLYMDFL 621
 Db 582 QMSYSISYVRGEGVILSYIGDREDLHCDIAETCFQKSEAILRELDVENAKLYLDFI 641
 QY 622 VLGIFFLAURLLALVLRVVKSR 646
 Db 642 VLGIFFISLRLAIYFLRYKRAER 666
 RESULT 4
 WHIT_DROME STANDARD; PRT; 687 AA.
 AC P10090; Q9V3A2; Q9Y333;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE White protein.
 GN W OR EG:BACN3B1.1 OR CG2759.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RX MEDLINE=90221897; PubMed=2109311;
 Pepling M., Mount S.M.;
 "Sequence of a cDNA from the Drosophila melanogaster white gene.";
 Nucleic Acids Res. 18:1633-1633(1990).
 [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=85134865; PubMed=6084717;
 O'Hare K., Murphy C., Lewis R., Rubin G.M.;
 "DNA sequence of the white locus of Drosophila melanogaster.";
 J. Mol. Biol. 180:437-455(1984).
 [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=21100348; PubMed=11156992;
 Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niwa S.,
 Yamamoto D.;
 "Dual-tagging gene trap of novel genes in Drosophila melanogaster.";
 Genetics 157:727-742(2001).
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.C., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 Brandon R.G., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster";
 Science 287:2185-2195(2000).
 [5]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Oregon-R;
 RC MEDLINE=20196011; PubMed=10731137;
 Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
 Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
 Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 Modolell J., Peter A., Schoettler P., Werner M., Mourikoti F.,
 Beinert D., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 Callister D.M., Campbell L.A., Darlamsou A., Henderson N.S.,
 McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 Glover D.M.;
 "From sequence to chromosome: the tip of the X chromosome of D.
 melanogaster";
 Science 287:2220-2222(2000).
 [6]
 RN SEQUENCE OF 224-331 FROM N.A.
 RX MEDLINE=89339145; PubMed=2503416;
 Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;
 "Cloning and characterization of the scarlet gene of Drosophila
 melanogaster";
 Genetics 122:595-606(1989).
 CC -1- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY
 FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS
 RESPONSIBLE FOR EYE COLOR. WHITE DIMERIZE WITH BROWN FOR THE
 TRANSPORT OF GUANINE AND WITH SCARLET FOR THE TRANSPORT OF
 TRYPTOPHAN.
 CC -1- SUBUNIT: HETERODIMER OF WHITE WITH EITHER BROWN OR SCARLET.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X51749; CAA36038.1; -
 DR EMBL; X02974; CAA26716.1; -

```

619 INQWADVEPGEISCTSSNTTCP--SSGKVILETLNFSRAADLPIDYVGLAILIVSFRVLAYL 677
637 VLRYRVK 643
678 ALRLRAR 684

RESULT 5
WHIT_CERCA STANDARD: PRT; 679 AA.
ID AC QI7320;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
DE W.
OS Ceratitiss capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Tephritidae; Tephritidae; Ceratitis.
OX NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RZ MEDLINE=96123276; PubMed=8533095;
RA Zwiabell L.J., Sacccone G., Zacharopoulou A., Besansky N.J.,
RA Evia G., Collins F.H., Louis C., Kafatos F.C.;
RT "The white gene of Ceratitiss capitata: a phenotypic marker for
RT germline transformation."
RL Science 270:2005-2007(1995).
CC -|- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89933; CAA61998.1; -.
DR InterPro; IPR003439; ABC_transportr.
DR DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF000005; ABC_tran; 1
DR ProDom; PD000006; ABC_transportr; 1.
DR TRIFRAMS; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSMEMBRANE; Transport.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 121 128 ATP (BY SIMILARITY).
FT TRANSMEM 427 445 POTENTIAL.
FT TRANSMEM 457 477 POTENTIAL.
FT TRANSMEM 507 525 POTENTIAL.
FT TRANSMEM 534 555 POTENTIAL.
FT TRANSMEM 568 586 POTENTIAL.
FT TRANSMEM 651 670 POTENTIAL.
FT CARBOHYD 628 628 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 679 AA; 75145 MW; 3F9CBC78A835C4CC CRC64;

Query Match 32.1%; Score 806; DB 1; Length 679;
Best Local Similarity 32.4%; Pred. No. 1.le-46;
Matches 194; Conservative 114; Mismatches 231; Indels 60; Gaps 14;

QY 82 KTLKCLSGKFCRRRELIGIMGPSACKSTFMNLAGYRESGMKG-QI-----LVNGRPR 134
DB 101 KULLKNDSGVAYPGELLAVNGSSGAGTKLLUNASA-FRSS-KGVQLSPSTRMLNGHPV 157
QY 135 ELRTRFKMSCYTIMODMLPLHLTVLEAMVMYSANLNUTENPDVRNDL--VTEILTALGLMS 192
DB 158 DAKEMAQRCAYVOODDLFTGSLTAREHLIFQAMVRPHMTOKORVDQVTDLSLGK 217

```



```
Qy 193 CSHT-----RTALLSGQRKRLAIAELVNNPVMFFDEPTSGDLSASCFOVVSMLKSL 246
Dy 218 QNTLIGVPGVKGLSGGRKRLAFASEALTDPLLCIDPTSGDLSFMAHVSQVVLKSL 277
Qy 247 AQGRFTICIHQPSAKLFEWFKLYLTLGOCIFKGVVNTLPYKGLGLHPTYNPA 306
Dy 278 SOKGTIVLTIHQPSSELFEKILLMAEGRVAFGLTGPGEAVDFSYIGATCPTNTPTA 337
Qy 307 DFIEVAS---GEYGLDNLPMFLRAVQGLCMAEKK-----SSPEKNEVPA 349
Dy 338 DFYVOVLAVVPGREVSRRVAKICDNFVAGKVSREMEQFKLVKSGNGEKEDENE--- 394
Qy 350 PCPPCPPEVDPIESHTFATSLTQFCILFKRTFLSILRDLVTLHLRPMHSHVIGVLGLL 409
Dy 395 -----YTKASWFMQFRAVLWRSWLSVLKPLLVKRLQLTTWAVLIGLI 440
Qy 410 YLHIGDDASK--VFNNTGCLFFSMLFLMFAALMPTVTLFLEMAVFMREHLNHYWLSKAY 467
Dy 441 FL--GOOLTOQVGMNNGAIFLFTNMTFQNSFATITVFTTELPVFMRETRSLRYCDTY 498
Qy 468 YLAKTMADVPFQVPCPVVYCVSVVMTGQPAETSRFLFSALATATALVAOSLGLLIGAA 527
Dy 499 FLGKTIAELPLVLPFLTAIAYPLIGLRPGVDHFFETALALVTLVANVSTSGYLISCA 558
Qy 528 SNSLOVATFVGPVTAIVPLVLSGFEVSKTIPTIYLOWSSYLSVYRVEGFIYIYG-ME 586
Dy 559 CSSTSMALSGPVVPIIPFLFGGFLNSGSPVYFKWLSVLSWFRYANEGLLINQWADVK 618
Qy 587 RGLDTCL--EERCPRFQSTLRALDVEDAKLYMDFLVGLGIFLALLLAYLVLYRVK 643
Dy 619 PGEITCLSNTPC--SSGEVILETLNFSASDLPDFDTGLALLIVGFRISAYIALTMAR 676

RESULT 6
WHIT_LUCCU
ID WHIT_LUCCU STANDARD; PRT; 677 AA.
AC Q05360;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Oestroidea; Calliphoridae; Lucilia.
RN NCBI_TaxID=7375;
RP
RX MEDLINE=97087158; PubMed=8933176;
RA Garcia R.L., Perkins H.D., Howells A.J.;
RT "The structure, sequence and developmental pattern of expression of
RT the white gene in the blowfly Lucilia cuprina.";
RL Insect Mol. Biol. 5:251-260(1996).
RN [2]
RP
RX MEDLINE=90264941; PubMed=1971656;
RA Elizur A., Vacek A.T., Howells A.J.;
RT "Cloning and characterization of the white and topaz eye color genes
RT from the sheep blowfly Lucilia cuprina.";
RL J. Mol. Evol. 30:347-358(1990).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
CC NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CC CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U38899; AAA2057.1; -
DR EMBL: X53265; CAA37365.1; -
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_Transport.
DR InterPro: IPR005284; Pigment_permease.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transport; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRfams: TIGR00955; 3a01204; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 119 126 ATP (POTENTIAL).
FT TRANSMEM 431 451 POTENTIAL.
FT TRANSMEM 456 476 POTENTIAL.
FT TRANSMEM 506 526 POTENTIAL.
FT TRANSMEM 534 554 POTENTIAL.
FT TRANSMEM 563 583 POTENTIAL.
FT TRANSMEM 647 667 POTENTIAL.
SQ SEQUENCE 677 AA; 75365 MW; D16FC11C97EED51D CRC64;

Query Match 23.7%; Score 794.5; DB 1; Length 677;
Best Local Similarity 31.7%; Pred. No. 6.7e-46;
Matches 192; Conservative 112; Mismatches 251; Indels 51; Gaps 12;

Qy 81 YKTLKLCISGKFCRR-----ELIGIMPSGAGSKFMILAGYR 119
Dy 77 WKQLNVNRKGVFCNERHIPRKHILKNVCQVAYPGELLAVMGSGAGKTLTLLNALAFRS 136
Qy 120 ESGMGQI-----LVNGRPRELTRFKMCSYIMQDDMLLPHLTVLEAMVSAANLTEN 173
Dy 137 ARGV--QISPSVRMLNGHPVDKEMQARCAVQDDDLFIGSLTAREHLIQAIVRMPT 194
Qy 174 PDVKNDL--VTEILTALGLMSCHT-----RTALLSGQRKRLAIALELVNNPVMFFD 225
Dy 195 MTQOKLQVRDQVIQDLSLIKCQNTIIGVPGRVKGLSGGRKRLAFASEALTDPLLCID 254
Qy 226 EPTSGDLSASCFOVVSMLKSLAOGRTTICTIHOPSAKLRFEMFKLYLTLGOCIFKGVV 285
Dy 255 EPTSGDLSFMAASVYQVLKLSQKQVILTIHQPSSELPFLDKILLMAGRVAFGLTGP 314
Qy 286 TNLPLYLKLGLHCHPTYNHPADFIIEVAS---GEYGLDNLPMFLRAVQGLCMAEKKSSP 342
Dy 315 VEAVDFSFIGAQCPTNYPADFYVQVLAVVPGREIESRDRISKCDN--FAVGKVSREM 372
Qy 343 EKNEVPAPCPPEVDPIESHTFATSLTQFCILFKRTFLSILRDLVTLHLRPMHSHVVI 402
Dy 373 EQNFQKIAAKTDGLQKDDETTILYKASFTQFRAIMWRSWISTLKEPLLVKRLVLIQTMTV 432
Qy 403 GVLIGLLYLHIGDDASKVFNNTGCLFSMLFLMFAALMPTVTLFPLEMAVFMREHLNHYW 462
Dy 433 AVLGLIFLQNPQMTQVGMNNGAIFLFTNMTFQNSFATITVFTTELPVFMRETRSLRY 492
Qy 463 SKAYILAKTMADVPFQVPCPVVYCVSVVMTGQPAETSRFLFSALATATALVAOSLGL 522
Dy 493 RCDTYFLGKTLAELPLFLVVPFLTAIAYPMIGLRPGITHTLSALALVTLVANVSTSGY 552
Qy 523 LIGAASNLQVATFVGPVTAIVPLVLSGFEVSKTIPTIYLOWSSYLSVYRVEGFIYI 582
Dy 553 LISCASTSMALSGVGPPLTIPTFLFGGVFLNSGSPVYFKWLSVLSWFRYANEGLLINQ 612
Qy 583 YG-MERGDLTCLER--CPFREPOSILRALDVEDAKLYMDFLVGLGIFLAL--RLAYLV 637
Dy 613 WADQPGEITCLSNTPC--SSGVXVLETLNFRDK---FTPLRYGLILLIIFRITAGVVA 668
Qy 638 LRYRVK 643
Dy 669 XKMRTK 674

RESULT 7
```

```

WHIT_ANOQA
ID WHIT_ANOQA STANDARD; PRT; 695 AA.
AC Q27256; Q17006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE white protein.
GN W.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Suakoko / G3;
RX MEDLINE=96423158; PubMed=8825759;
RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfler D.,
RT Collins F.H.;
RT "Cloning and characterization of the white gene from Anopheles
gambiae".
Insect Mol. Biol. 4:217-231(1995).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U29486; AAC46995.1; -
CC EMBL; U29485; AAC46994.1; -
CC EMBL; U29484; AAC47423.1; -
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transportr.
CC ProDom; PD000005; Pigment_permease.
CC Pfam; PF00005; ABC_tran; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRFAMS; TIGR00955; 3a01204; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 133 140 ATP (POTENTIAL).
FT NP_BIND 288 295 ATP (POTENTIAL).
FT TRANSMEM 444 464 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT TRANSMEM 524 544 POTENTIAL.
FT TRANSMEM 552 572 POTENTIAL.
FT TRANSMEM 581 601 POTENTIAL.
FT TRANSMEM 669 689 POTENTIAL.
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 100 100 N -> S (IN REF. 1; AAC47423).
FT CONFLICT 691 693 SRS -> YAR (IN REF. 1; AAC47423).
SQ SEQUENCE 695 AA; 77218 MW; EE8B9517239B2961 CRC64;

Query Match
Best Local Similarity 31.7%; Pred. No. 7.6e-45;
Matches 204; Conservative 117; Mismatches 276; Indels 46; Gaps 14;

Qy 40 VENHITEAQRFSHLPKRSADVIEFE--LSYSVREGPCWRKRK-----YKTLK 86
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 VKGIPOCERLTWTWKE--IDV-FGEAPTQDKPREPLCTRLRNCCTRQRKDFNPRKHLK 117
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 87 CLSKKCFRRELIGMGPSGAGKSTFMILAGYRESGK-----GQILVNGRPRLRFRKM 142
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 NVTVGAKSGELLAVMGSSGAGKTTLLNALAFRSPPGVKISPNVAVNGVPVNAEQLRAR 177
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8

WHIT_ANOAL

```

ID WHIT_ANOAL STANDARD; PRT; 709 AA.
AC Q16928;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Anopheles albimanus (New world malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Anopheles.
OX NCBI_TaxID=7167;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Teco;
RX MEDLINE=98126549; PubMed=9465401;
RA Ke Z., Benedict M.Q., Cornel A.J., Besansky N.J., Collins F.H.;
RT "The Anopheles albimanus white gene: molecular characterization of
RT the gene and a spontaneous white gene mutation.";
RL Genetica 101:87-96(1997).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```
CC EMBL: L76302; AAA88240.1; ..
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF0005284; Pigment_permease.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR TIGRFAMs: TIGR00955; 3a01204; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 136 143
FT ATP (POTENTIAL).
FT NP_BIND 292 299
FT ATP (POTENTIAL).
FT TRANSMEM 457 475
FT POTENTIAL.
FT TRANSMEM 487 507
FT POTENTIAL.
FT TRANSMEM 537 555
FT POTENTIAL.
FT TRANSMEM 564 585
FT POTENTIAL.
FT TRANSMEM 598 616
FT POTENTIAL.
FT TRANSMEM 681 700
FT POTENTIAL.
FT CARBOHYD 485 485
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 658 658
FT N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 709 AA; 79052 MW; FBD76D5C69D3BECA CRC64;

Query Match 22.0%; Score 736.5; DB 1; Length 709;
Best Local Similarity 32.4%; Pred. No. 5.4e-42;
Matches 199; Conservative 97; Mismatches 244; Indels 75; Gaps 13;

QY 82 KTLKCLSGKFCRRELIGIMSGAGKSTFMNLAGYRESGMK----GQILVNGRPRELR 137
DB 116 KHLKNVTGVARSGELLAVMGSSGAGKTTLLNELAFRSPGVKISPNAIRTLNGVPVTA 175
QY 138 TPRKSCYIMQDDMLPHLTIVLEAMVSNANLTEN---PDVKNDLVTEILTALGLMSCS 194
DB 176 QMRACAYVQDDLFPSLTTEKHLMFQMLRGMGRDVPATPIKMRHVDEVLOELSLVKA 235
QY 195 HT-----RTALLSGQRKRLAIALELVNPNPMFFDEPTSGDLSASCFOVYSLMKSLAQ 248
DB 236 DTIIIGVAGRVKGLSGERKRTAFRSETLTDPHLLCDEPTSSLDSEMAQSVLQVLKGMAM 295
QY 249 GORTIICITHOPSAKLFEMFKLYLISQGCQIFKGVVNTLIPYKGLGLHCHCTYNPAPDF 308
DB 296 KGKTIILTHOPSSSELYCLFDRILLVAEG-VAFLGSPYQASADFFSGLGICPCPNYNPAPDF 354
QY 309 -----IIIVASGEYGDNLNPNLFRVQNLCA 334
DB 355 YVQMLAIAPNKECTEGETIKKICDSFAVSPAIARDIETASVNGD-----GGTE 403
QY 335 MAEKKSSPEKNEVPAPCPPEVDPPIESHTSTATLTQF-CILFKRTLSILRTVLT 393
DB 404 LTRTKHTD-----PYFLQPMGEVDSTGYRASWWTQFYCILW-RSLVSLVKDPLVK 454
QY 394 LRFMSHVIVGLIGLLYHIGDDASKVFNNTCGLFFSMLFLMFAALMPTVLTFLPMAYF 453
DB 455 VRLQATAMVASLIGSIYFGQVLDQGVNMINGSLFLTNMTFQNVFAVINVESAEPLVF 514
QY 454 MEHLNLYMSKAYLAKTMADVPQVCPVYVCSTVYVMTQPAETSRFLFSALATAT 513
DB 515 LREKRSRLRVDTYFLGKTIABELPLFIAPVFTSTIYPMIGLKAISHYLTFLITVL 574
QY 514 ALVAQSLGLLGAANSLOVATFVGPVTAIPVLLSGFEVSEFKTIPTYLQWSSYLSYVR 573
DB 575 ANVSTFGLYLSCASSISMAISVGPVVPVPIFLFGGFLNSASPAYFKYLSYLSWFR 634
QY 574 GREGVILTYGMR-GDLTCLER--CPPREQSLRALD--VEDAKLYMDFLVIGIFPL 628
DB 635 ANEALLINOWADHRDGEIGCTRANVTCP-ASGEIILETFNFRVEDFALDGLCFALIVLF 693
QY 629 ALRLLAYLVLYRVK 643
DB 694 RLGAFLCLMLRSRK 708

RESULT 9
YPC3_CAEEL
ID YPC3_CAEEL STANDARD; PRT; 598 AA.
```

```
AC Q11180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative ABC transporter C05D10.3 in chromosome III.
GN C05D10.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: U13645; AAA20989.2;
DR WormPep; C05D10.3; CE29170.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
KW Hypothetical protein; ATP-binding; Transmembrane; Transport.
FT NP_BIND 27 34
FT ATP (POTENTIAL).
FT TRANSMEM 336 356
FT POTENTIAL.
FT TRANSMEM 425 445
FT POTENTIAL.
FT TRANSMEM 453 473
FT POTENTIAL.
FT TRANSMEM 478 498
FT POTENTIAL.
SQ SEQUENCE 598 AA; 66906 MW; 9D6414E06898E343 CRC64;

Query Match 21.6%; Score 723.5; DB 1; Length 598;
Best Local Similarity 28.4%; Pred. No. 3.3e-41;
Matches 171; Conservative 126; Mismatches 255; Indels 51; Gaps 9;

QY 78 KRGYKTLKCLSGKFCRRELIGIMSGAGKSTFMNLAGYRESGM--KQILVNGRPRE 135
DB 3 KRKVEILHNVSQMAESGKLLAILGSSGAGKTTLMNVLTSRLNLDVQSGILIDRRAN 62
QY 136 LRTFRKMSYIMQDDMLPHLTIVLEAMVSNANLTEN---PDVKNDL-VTEILTALGLMS 192
DB 63 KWKIREMSAFVQDHDMFVCTMTAREHLQFMARLMDGDQYVSDHERQLRVEQVLTOMGLK 122
QY 193 CSHTRTAL-----LSGGQRKRLAIALELVNPNPMFFDEPTSGDLSASCFOVYSLMKSL 246
DB 123 CADTVIGIPNQLKGLSCGEKRLSFASEILTCKPLFCDEPTSGLDAMFAGHVVAQLRSL 182
QY 247 AGGGRITICTIHOPSAKLFEMFKLYLISQGCQIFKGVVNTLIPYKGLGLHCHCTYVHNP 306
DB 183 ADNGMTVIITIHQPSHHVYSLFNVCMLMACGRVIYILGPDQVAPLFEKGCYPCPAYNFA 242
QY 307 DFIEI-----VASGEYGDNLNPNLFRAV-----ONGICAMAEKSSPEKNE 346
DB 243 DHLIRTLAVIDSDRATSMKTIKIRQGFSLTDLQSVLAIGNANKLRAASFVTSQDTSEK 302
QY 347 VPAPCPPEVDPPIESHTSTATLTQFCILFKRTLSILRTVLTFLHFMHVIVGLI 406
```


	ENBL; X59720; CAA42328.1;	-	
DR	PIR; S19421; S19421.		
DR	PIR; S40914; S40914.		
DR	SGB; S000604; ADP1.		
DR	InterPro; IPR003593; AAA_ATPase.		
DR	InterPro; IPR003439; ABC_transportr.		
DR	Pfam; PF00005; ABC_tran; 1.		
DR	Frodcom; PD000006; ABC_transportr; 1.		
DR	SMART; SM00382; AAA; 1.		
DR	PROSITE; PS00211; ABC_TRANSPORTER; 1.		
KW	ATP-binding; Transmembrane; Glycoprotein; Signal.		
FT	SIGNAL	1	25 POTENTIAL.
FT	CHAIN	25	1049 PROBABLE ATP-DEPENDENT PERMEASE.
FT	NP_BIND	423	430 ATP (BY SIMILARITY).
FT	TRANSMEM	325	345 POTENTIAL.
FT	TRANSMEM	464	481 POTENTIAL.
FT	TRANSMEM	794	814 POTENTIAL.
FT	TRANSMEM	829	849 POTENTIAL.
FT	TRANSMEM	878	898 POTENTIAL.
FT	TRANSMEM	910	930 POTENTIAL.
FT	TRANSMEM	938	958 POTENTIAL.
FT	TRANSMEM	1001	1021 POTENTIAL.
FT	TRANSMEM	1025	1045 POTENTIAL.
FT	CARBOHYD	50	50 N-LINKED (GLCNAC. .) (POTENTIAL).

FT	CARBOHYD	114	114	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	165	165	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	815	815	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	935	935	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	960	960	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	971	971	N-LINKED (GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	1049	AA; 117231 MW; ABC9CE54BCDF6A3 CRC64;		

Query Match		17 88;	Score 595;	DB 1;	Length 1049;
Best Local Similarity		25.18;	Pred. No. 2.5e-32;		
Matches 181;	Conservative 140;	Mismatches 239;	Indels 160;	Gaps 20;	
Qy	53 LPKRSVD-----	IEEVELSYSYREGPCWRKRGY-KTLKCLSGKFCRRELIG	99		
Db	364 LPDEDAVNFQNEDDTFLATLSFNITYSV---PSINDGVEETVLNEISGIVKPGQILA	420			
Qy	100 IMGPSGAGKTFWNILAGYRESG-MKGQILVNGRPRELRTFRKMSCYIMODMLLPHLTV	158			
Db	421 IMGSGAGKTTLLDILLAMKRKTGHVSGIKVNGISMDRKSFSKIIGFVDQDDFLLTTLTV	480			
Qy	159 LEAMMYSANLNLTE--NPQVKNLDVTEILTALGIMS-----CSHTRTALLSGQCKRLAI	211			
Db	481 FETVLNSALLURPKALSFCAKARVYKVLEELRIIDIKRIIENEFDRGISGEGEKRRVSI	540			
Qy	212 ALELVNPPVMFDEPTSGDLSASCQVQVSLMKSLAGQ-CRTIICITHOPSAKLFEMFDK	270			
Db	541 ACELVTSPLVLFDEPTSGDLSANANNVTECLVRLSSDYNRTVLSHIQPRSNIFYLFDK	600			
Qy	271 LYLSQGQCIFKGVNVLNLIPLYKGLGHLCHPTIYNPADFIEVA--SGEYGDLPMLFRAV	328			
Db	601 LVLLSKGEMVYSGNAKKVSEFLRNEGICYPDNYNIADYLDITFEAGPQGR-----RRI	655			
Qy	329 QNGLCAEAKKSPENKVPAPCPGCPPEVDPTESHTFATST-----	370			
Db	656 RN----ISDLEAGTDTNDID-----NTIHQTFTSSDGTQREWAHLAAHRDEIR	701			
Qy	371 -----	-----	370		
Db	702 SLLNDEEDVGTGRRGATEIDINTLLHDKYKDSVYVYAELSOEIEEVLSEGEESNVLN	761			
Qy	371 -----LTQFCILFKRTFLSLRDTVLTHLRFMSHVYVIGVLIGLLYLIHGDDAS	418			
Db	762 GDLPTGQOSAGFLQQLSILNSRSEFKMYRNPKLLGNLYLTLLSLFLGTLTYVNSDTS	821			
Qy	419 KVENNTGCLFFSMLFLMFAALMTVTLFPLEMAVFMREHLNYYYSKAYLAKTMAD-VP	477			
Db	822 GFQNRMGFLPFILTYFGFVT-FTGLSSFALERIIFIKERSNNYSPPLAYYISKINSEVP	880			
Qy	478 FOVVCPPVYCSIVYVMTGQPAETSFR-----LLFSALATATLVAQSLGLLIGAANS	530			
Db	881 LRVVPPILLSLIYVPMGLNMKNNAKFKIGIILFN---LGISLEILTIIGIIFEDLNS	937			
Qy	531 LOVATFVGPVTAIPVLLFSGFFVSKFTIPTY-LOWSSYLSYVRYGFEGLTIYGMERG	589			
Db	938 IILSVLV-----LLGSLFLFSLGFLNTKNITNVAFKLNFSVYAYESLLIN---EVKT	989			
Qy	590 LTLCEKRCPPR---EPQSIILRALDVEDAKLYMDFVLGIGTFLLRALLAVLYLRVYKSR	646			
Db	990 LMLKERRYGLNIEVPGATILSTFGVYVQNLVDFIKILALFNVVFLIMGYLAKWIVVEQR	1049			

RESULT 13			
ABG5_HUMAN	STANDARD;	PTT;	651 AA.
ID	ABG5_HUMAN		
AC	Q9H222;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	ATP-binding cassette, sub-family G, member 5 (Sterolin-1).		
GN	ABCG5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=20578753; PubMed=11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 RA Dean M., Patel S.B.;
 RA "Identification of a gene, ABCG5, important in the regulation of
 RT dietary cholesterol absorption."
 RL Nat. Genet. 27:79-83(2001).
 RN [2]
 RP TISSUE SPECIFICITY, AND INDUCTION.
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;
 RA "Accumulation of dietary cholesterol in sitosterolemia caused by
 mutations in adjacent ABC transporters."
 RL Science 290:1771-1775(2000).
 CC -!- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower
 CC level, in the liver.
 CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated
 CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF312713; AAG53097.1; -;
 DR MGD; MGI:1351659; Abcg5.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transportr.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transportr; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT DOMAIN 1 385
 FT TRANSMEM 386 406
 FT DOMAIN 407 422
 FT TRANSMEM 423 443
 FT DOMAIN 444 463
 FT TRANSMEM 464 484
 FT DOMAIN 485 504
 FT TRANSMEM 505 525
 FT DOMAIN 526 529
 FT TRANSMEM 530 550
 FT DOMAIN 551 622
 FT TRANSMEM 623 643
 FT DOMAIN 644 652
 FT NP_BIND 87 94
 FT CARBOHYD 410 410
 FT CARBOHYD 585 585
 FT CARBOHYD 592 592
 SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match

17.3%; Score 577.5; DB 1; Length 652;

Best Local Similarity 26.3%; Pred. No. 2.2e-31;
 Matches 176; Conservative 137; Mismatches 272; Indels 83; Gaps 19;
 QY 26 DGAEPPTLTHLKKVENHTEAORPSHLPKSAVDIEFVELSYSVREGPCW-----RRRG 80
 DB 10 EGARGP-----HINRGSLSSLEQGSVGTGEARHSLGLVHVSYSVSNRVGPMWNIKSCQKW 65
 QY 81 YKTLKCLSGKFCRRELIGINGPGAGKSTFEMNLAG-YRESG-MKGQILVNGRPRELRT 138
 DB 66 DRQILKDVSLYIESGQIMCIIGSSGSGKTTLLDAISGRRLRTGTLEGEVFNNG--CELR 123
 QY 139 FRKMSC--YIMQDDMLLPHLTLEAMVYSANLNL--TENPDVKNLDLVLTILTALGLMSCS 195
 DB 124 DQFQDCFSYQLQSDVFLSLVRETLRYTAMALCRSSADFYNNKVEAVMTLSL---SH 180
 QY 196 TRTAL-----LSGGORKRLATALELVNPPVNFDEPTSGLDSDSCFQVSVLMKSLA 247
 DB 181 VADQMGISYNGFGISSGERRRVSIAAQLQDPKVMMLDEPTGLDCMTANQIVLLLAELA 240
 QY 248 QGGRTIICITHOPSAKLFEMFDKLYLSQGCIFKGVVNTLPIYKGLGLHCPTYHNPAD 307
 DB 241 RDRIVIVITHQPRSELEFHQDKIALLTYGELVFCGTPPEMLGFENNCGYPCPEHSNPF 300
 QY 308 FIEVASGEYGDLPML--FRAVQNGLCMAEKK-----SSPEKNEVPAPCPGCPPEV- 358
 DB 301 FYMDLTSVDTQSREREIETRYKRVQMLECAFKESEDYHKKILENIERYARKLTLPVFPKTK 360
 QY 359 DPESHTTATSTLTQFCILFKRTFLSILRDTVLTHLRPMHSHVVGVLGLLGLHIGDDAS 418
 DB 361 DP-----PGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLMGLFIFYLLRVQNNTL 413
 QY 419 K--VFNNTGCLFFSMLFLMFAALMPTVLTPLPMAVEMREHNLVWYSLKAYLAKTMADV 476
 DB 414 KGAVQDRVGLLYQLVYGATPYTGMLNANLFPMLRAYSDQESQDGLYHKWQMLLAYVLVL 473
 QY 477 PFQVVCPPVYCSIVYMMTGQPAETSRFLFLFSALATATALVAQSLGL-LIGAASNSLOVAT 535
 DB 474 PFSVIATVIFSVCYMTGLYPEVARFGYFSAALLAPHLIGEFLLVLGLIVQNPINVS 533
 QY 536 FVGPTAIPVLLFSGFFVFSFKTIPTYLQWSSYLSVYRVGFEVILT-IYMERGDLTC-- 592
 DB 534 IVALLSISGLLIGSGFIRNIQEMPPLKILGYTFQKCCETLVVNEFYGL---NFTCGG 590
 QY 593 -----LEERCPRFPOSILRALDVEDAKLYMDFVLVGLGIFFLALRL 633
 DB 591 SNTSMNLNHPMCAITOGVQFIEKTCP-----GATSRFTANFLILYGFIALVIL 638
 QY 634 AYLVLYR 641
 DB 639 GIVIFKVR 646

Search completed: June 10, 2003, 16:46:28
 Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:38:29 ; Search time 44 Seconds
(without alignments)
1411.428 Million cell updates/sec

Title: US-10-072-621-9
Perfect score: 3347
Sequence: 1 MAEKALEAVCGGLPGAVAM.....FLALRLAYLVLRVRSER 646
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

rched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3309	98.9	646	2 JC7777	ATP binding cassette
2	2350	70.2	638	2 G02068	white homolog - hu
3	2161	64.6	416	2 T46401	hypothetical prote
4	879.5	26.3	687	1 FYFW	white protein - fr
5	760.5	22.7	646	2 C86441	probable ABC trans
6	716.5	21.4	658	2 T31958	hypothetical prote
7	699.5	20.9	609	2 E96742	probable ABC trans
8	697	20.8	737	2 T46101	ABC transporter-li
9	686	20.5	695	2 T21109	hypothetical prote
10	685	20.5	559	2 B88474	protein C0SD10.3
11	665.5	19.9	725	2 C84423	probable ABC trans
12	663.5	19.8	590	2 B96573	protein F12M16.17
13	640	19.1	635	2 T08934	hypothetical prote
14	635.5	19.0	608	2 T34391	hypothetical prote
15	626	18.7	577	2 T04229	ABC-type transport
16	613	18.3	649	2 A84509	probable ABC trans
17	612	18.3	725	2 T47652	ABC transporter-li
18	611	18.3	678	2 H96552	hypothetical prote
19	607	18.1	633	2 T19189	hypothetical prote
20	600	17.9	639	2 G88939	protein C10C6.5
21	598.5	17.9	720	2 T47648	ABC transporter-li
22	596	17.8	687	2 D96553	hypothetical prote
23	595	17.8	1049	1 S19421	ATP-dependent perm
24	586.5	17.5	739	2 T45891	ABC transporter-li
25	585.5	17.5	547	2 T31543	hypothetical prote
26	585	17.5	740	1 T02567	probable ATP-bindi
27	584.5	17.5	1435	2 D96693	protein Putative A
28	584	17.4	708	2 T47650	ABC transporter-li
29	577	17.2	610	2 T19353	hypothetical prote

ALIGNMENTS

RESULT 1
JC7777

ATP binding cassette transporter, subclass G4 - human

C:Species: Homo sapiens (man)

C:Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 02-Apr-2002

C:Accession: JC7777

R:Engel, T.; Lorkowski, S.; Lueken, A.; Rust, S.; Schlueter, B.; Berger, G.; Cullen,

Biochem. Biophys. Res. Commun. 288, 483-488, 2001

A:Title: The human ABCG4 gene is regulated by oxysterols and retinoids in monocyte-de

A:Reference number: JC7777

A:Contents: Macrophages

A:Accession: JC7777

A:Molecule type: mRNA

A:Residues: 1-646 <ENG>

A:Cross-references: GB:AJ308237

C:Comment: This protein, a novel member of the ATP binding cassette(ABC) half-size tr

C:Genetics:

A:Gene: abcg4

A:Map position: 11q23.3

C:Keywords: macrophage

Query Match 98.9%; Score 3309; DB 2; Length 646;

Best Local Similarity 98.9%; Pred. No. 5.2e-251;

Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY	1	MAEKALEAVCGGLPGAVAMAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSVD	60
DB	1	MAEKALEAVCGGLPGAVAMAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSVD	60
QY	61	IEFVELSYSVREGPCWKRKGYKTLKCLSGKFCRRRELIGIMGPGACKSTFMNIIAGYRE	120
DB	61	IEFVELSYSVREGPCWKRKGYKTLKCLSGKFCRRRELIGIMGPGACKSTFMNIIAGYRE	120
QY	121	SGMKGOILVNGRPRELTFKMSYIMQDDMLLPHTLVLEAMVSNANLNTENPDVKNDL	180
DB	121	SGMKGOILVNGRPRELTFKMSYIMQDDMLLPHTLVLEAMVSNANLNTENPDVKNDL	180
QY	181	VTEILTALGLMSCSHRTALLSGGQRRLAIALELVNPPVPMFDEFTSLGDSASCFQV	240
DB	181	VTEILTALGLMSCSHRTALLSGGQRRLAIALELVNPPVPMFDEFTSLGDSASCFQV	240
QY	241	SLMKSLAQQGRTICTIHQPSAKLFEMFDKLYILSQQCIFKGVVNTNLIPLYKGLGLHCP	300
DB	241	SLMKSLAQQGRTICTIHQPSAKLFEMFDKLYILSQQCIFKGVVNTNLIPLYKGLGLHCP	300
QY	301	TYHNPAADFTIEVASGEYGLDNLPMFLRAVQNGLCMAEKKSSPEKNEVPAPCPPEVDP	360
DB	301	TYHNPAADFTIEVASGEYGLDNLPMFLRAVQNGLCMAEKKSSPEKNEVPAPCPPEVDP	360
QY	361	IESHTATSTLTQFCILFKRTFLSILRDTVTLHLRFNSHVIVGLIGLLYHIGDDASKV	420
DB	361	IESHTATSTLTQFCILFKRTFLSILRDTVTLHLRFNSHVIVGLIGLLYHIGDDASKV	420

Query Match	70.2%	Score	2350;	DB 2;	Length	638;			
Best Local Similarity	69.9%	Pred.	No. 6.3e-176;						
Matches	453;	Conservative	80;	Mismatches	77;	Indels	38;	Gaps	4;

Qy	26	DGAEPPLVTHLKKVENHTEAORFSLPKRSADVIEFVLSVSRGPGCQWRKGYKTL	85
Db	2	EATETDLNLGHLLKVDNLTTEAORFSSLPRAAVNIEFRLSVSPGPMWRKGYKTL	61
Qy	86	KCLSGKFCRRELIGIMPGSAGKSTFNMLIAGYRESGMKGQILVNGRPRELRTFKM	145
Db	62	KGISGRFNSGELVAIMGPSGAGKSTLMLIAGYRETMKGAVLINGLPLDLRCFRK	121
Qy	146	IMODDMLLPHLTVLEAMVVSANLNTENPDVKNDLVTEILTALGLMSCSHTRTALL	205
Db	122	IMODDMLLPHLTVOEAMVVAHLLKQEKDGRREMKIEILTALGLLSCANTRTG	181
Qy	206	RKLRLATALELVNPPVMFFDEPTSGLDSASCFQVWSLMKSLAOGRTIICTIH	265
Db	182	RKLRLATALELVNPPVMFFDEPTSGLDSASCFQVWSLMKSLAOGRTIICTIH	241
Qy	266	EMFDKLYLSQGOCIFKGVVYTNLIPYLKGLGHCPTVHPNADFTIEVASGEY	325
Db	242	ELFDQLYLSQGOCVYRGKVCNLVPYLRDLGLNCPVTHNPADFVMEVASGEY	301
Qy	326	RAYQNGLCMAEAKKSPEKNEVPACPCPEVDPI-----	361
Db	302	RAYREGM-----SDSHKLDGDA-----EYNPFLWHRPSEVVKTKRLK	350
Qy	362	--ESHFTFATSTLQFCILFKRTFLSLRDLVTLHFRMSHWVGLTGLYLGHD	419
Db	351	MECHHSFASCLTQFCILFKRTFLSINRDSVLHRLTSHIGITGLLGLYLG	410
Qy	420	VFNNTGCLFSMLFLMFAALPMPTVLTFFLEMAVFMREHLNWSLKAYYLAK	479

Query Match	64.6%	Score 2161;	DB 2;	Length 416;
Best Local Similarity	100.0%;	Prod. No. 2.3e-161;		
Matches 416;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	231	LDSASCFQVWSLMSLAQGGRTIICTIHOPSAKLFEMEDKLYILSGOGCIFKGVVYNLIP	290	
Db	1	LDSASCFQVWSLMSLAQGGRTIICTIHOPSAKLFEMEDKLYILSGOGCIFKGVVYNLIP	60	
QY	291	YLGGLGHLCHPTYHNPADFIIEVASGEYGDLPNMLFRAVONGICLMAEKKSSPEKNEVPAP	350	
Db	61	YLGGLGHLCHPTYHNPADFIIEVASGEYGDLPNMLFRAVONGICLMAEKKSSPEKNEVPAP	120	
QY	351	CPPCPPEVDPTESHTFATSTLUTQCIILFKRFTLSLTRDTVLTHLFEMSHVTVGLVIGLGLY	410	
Db	121	CPPCPPEVDPTESHTFATSTLUTQCIILFKRFTLSILRDTVLTHLFEMSHVTVGLVIGLGLY	180	
QY	411	LHIGDDASKVFNNTGCLFFSMLFLMFMAALMPTVLTFFPLEMAVFMREHLNINWYSLKAYLYA	470	
Db	181	LHIGDDASKVFNNTGCLFFSMLFLMFMAALMPTVLTFFPLEMAVFMREHLNINWYSLKAYLYA	240	
QY	471	KTMAADVFPQVVCVYVYCSIVVYMTGQPAETSRFLFLFSALATATALVAOSLGLLIGAASNS	530	
Db	241	KTMAADVFPQVVCVYVYCSIVVYMTGQPAETSRFLFLFSALATATALVAOSLGLLIGAASNS	300	
QY	531	LOVATFVGVPVTAIPVLLFSGFFVFSFKTIPTPTQLQWSSYLSVYRGPGEVILTIYGMERGDL	590	
Db	301	LOVATFVGVPVTAIPVLLFSGFFVFSFKTIPTPTQLQWSSYLSVYRGPGEVILTIYGMERGDL	360	
QY	591	TCLEERCPPFREPOQIILRALDVEDAKLYMDFVLVLGFIFFLALRLLAYLVLYRVRKSSR	646	
Db	361	TCLEERCPPFREPOQIILRALDVEDAKLYMDFVLVLGFIFFLALRLLAYLVLYRVRKSSR	416	

RESULT 4
FYFW
white protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:date: 31-Dec-1990 #sequence_revision 17-Feb-1995
C:Accession: S08635; S07263; S10240

R:Pepling, M.; Mount, S.M.
 Nucleic Acids Res. 18, 1633, 1990
 A:Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.
 A:Reference number: S08635; MUID:90221897; PMID:2109311
 A:Accession: S08635
 A:Molecule type: mRNA
 A:Residues: 1-687 <PEP>
 A:Cross-references: EMBL:X51749; NID:g8825; PIDN:CAA36038.1; PID:g8826
 R:O'Hare, K.; Murphy, C.; Levis, R.; Rubin, G.M.
 J. Mol. Biol. 180, 437-455, 1984
 A:Title: DNA sequence of the white locus of *Drosophila melanogaster*.
 A:Reference number: S07263; MUID:85134865; PMID:6084717
 A:Accession: S07263
 A:Molecule type: DNA
 A:Residues: 1-24, 'LIFEIPYHCRVTAD', 30-334, 'ITLHNSYPAAWPSVLTPIRTTYRCWCLPCDGRSSPVIGSPRY
 A:Cross-references: EMBL:X02974
 A:Experimental source: strain Canton S
 O'Hare, K.
 Submitted to the EMBL Data Library, June 1995
 Reference number: S10240
 A:Accession: S10240
 A:Molecule type: DNA
 A:Residues: 1-24, 'LIFEIPYHCRVTAD', 30-687 <OHA2>
 A:Cross-references: EMBL:X02974; NID:g10873; PIDN:CAA26716.1; PID:g10874
 A:Experimental source: strain Canton S
 C:Genetics:
 A:Gene: white; w
 A:Cross-references: FlyBase:FBgn0003996
 A:Introns: 24/3; 116/1; 334/2; 439/3; 483/3
 A:Superfamily: fruit fly white protein; ATP-binding cassette homology
 C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
 F:113-317/Domain: ATP-binding cassette homology <ABC>
 F:130-137/Region: nucleotide-binding motif A (P-loop)
 F:261-265/Region: nucleotide-binding motif B
 F:67, 93, 472, 554, 651/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 26.3%; Score 879.5; DB 1; Length 687;
 Best Local Similarity 33.9%; Pred. No. 9.6e-61;
 Matches 206; Conservative 121; Mismatches 237; Indels 43; Gaps 12;

QY 73 GPCWRK-----RCY-----KTLKCLSGKCRRELIGIMPSGAGKSTFNNILA 116
 DB 85 GSGWQLVNTRGFLNCRHHPAPRHLKKNVCGVAYPGELLAVMGSSGAGKTTLLNALA 144
 QY 117 GYRESGKK-----GQILVNGRPRLRTFRKMSCYIMQDDMLLPHLTVLEAMVMSANLNLE 172
 DB 145 FRSPQGIQVSPSGMRLNGQVDAKEMQARCAVYQDDLLFGISLTAREHLIFQAVRMPR 204
 QY 173 NPDKNDL--VTEILTALGLMSCHT-----RTALLSGQKRLATALELVNPPVWF 224
 DB 205 HLTYRQVARVDQVIOELSLKCOHTIIGVGRVKGSLGGERKRLAFASEALDPPLLIC 264
 QY 225 DEPTSGLDASCFQVYSLMKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGCIFKG 284
 DB 265 DEPTSGLDSTAHVSQVQLKLSOKGKTIVLTTHQPSSELFDFKILLMAEGRVAFGLT 324
 QY 285 VTNLIPYKGLGLHCPTNHPADFIIEVAS---GEYGLNPMFLFRAVONGLCAMAEKSS 341
 DB 325 PSEAVDFSVYGAQCPTNYPADFYQVLAVYFGRREIESDRKACIDN----FATSKVA 380
 QY 342 PERNEVPAPCPPEVDPTESHTFATSTLTQCIILFKRTFLSILRDTVLTHLRFMSHV 401
 DB 381 RDMEQLLATKNLEKLEQPENGYTYKATPMQFRAVLWRSWLSVLEPLLVKVLQTTM 440
 QY 402 IGVILGLYLHIGDASK--VFNTGCLFFSMFLMFAALMPVLTLPPELMAVPMREHLN 459
 DB 441 VAILIGLIFL--GQULTQVGMINGAIFLFLNMTFQNFVATINVTSELPMREARS 498
 QY 460 YWISLAKYLAKTMAVDFQVCPVYVCSIVYWMTPGAPETSRLFFLSALATATALVAQS 519
 DB 499 RLYRCDTYFLGKTIAELPLELTVPVLTATAYPMIGLRAGVHLFFNCLALVTLVANVST 558
 QY 520 LGLLIGAANSLSQVATFVGVPVTAIPVLLSGFFVFSEKTIPTYLOWSSYLSYVRGFEV 579

DB 559 FGYLISCASSTSMALSVGPPVPIIPFLLEGFFLNSGSPVYLKWLKLSYLSWERYANEGL 618
 QY 580 LTIYG-MERGDLTCLER--CPFREPOSILRALDVEDAKLYMDFLVIGLFFLALRLAYL 636
 DB 619 INQWADVEPGEISCTSSNTTCC--SSQVILETLUNFSAADPLDYVGLAILIVSRVAYL 677
 QY 637 VLRYRVK 643
 DB 678 ALRLRAR 684

RESULT 5
 C86441
 probable ABC transporter [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86441
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C86441
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-646 <STO>
 A:Cross-references: GB:AE005172; NID:g1136734; PIDN:AAG31315.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F1216.1; AT

Query Match 22.7%; Score 760.5; DB 2; Length 646;
 Best Local Similarity 30.3%; Pred. No. 1.9e-51;
 Matches 189; Conservative 133; Mismatches 239; Indels 63; Gaps 19;

QY 57 SAVDIEFVELSVSR---EGPC---WRKRGYKTLKCLSGKCFRRELIGIMPSGAGKST 110
 DB 47 SMYPITLKEVYKVKTEQTSQCMGSKK-EKTLNGITGWCVPGEFLAMLPFGSGSKTT 105
 QY 111 FMILAGYRESGKMGQILVNGRPRLRTFRKMSCYIMQDDMLLPHLTVLEAMVMSANLN 170
 DB 106 LLSALGRUSKTFSGKVMYNGQPFSGCTIKRRTGFVAQDDVLPHTLVETLFTALLRL 164
 QY 171 TEN--PDVKNDLVTEILTALGLMSCSHRTVA----LLSGQKRLAIALELVNPPVMF 223
 DB 165 PSSILRDEKAHVDVRIAEGLNRCNSMIGGLPFRGISGGEKKRYSIGEMLINPSLLL 224
 QY 224 FDEPTSGLDASCFQVYSLMKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGCQIFKG 283
 DB 225 LDEPTSGLDSTTAHRIVTIKRLASGRTVVVTHQPSRIYHMFDRKLVLLSEGSPIYG 284
 QY 284 VTNLIPYKGLGLHCPTNHPADFIIEVAS---EYGLNPMFLFRAVONGLCAMAEKK 339
 DB 285 AASAEYFSSGLSTSLAVNPADLLDLANGIPPDQTKETSEQEQTKVETLVSAYEK 344
 QY 340 SSPEKNEVPAPCPPEVDPIESHNF----ATSTL--TQCFILFKRTFLSILRDTV-- 390
 DB 345 ISRK-----LKAELCNAESHSEYTKAAKNLKSQWCTTWYQVTLVLRQVRE 394
 QY 391 -----LTHLRFMSHVIGVLIGLHLIGDDASKVFNNNTCCLFFSMLFLMFAALMPTVLT 445
 DB 395 RFESEFNKLRIFQVISVAFGLGLLWH--TPKSHIQDRTALLFFFSVFWGFPYLVNAVFT 452
 QY 446 FPLEMAVFMREHLNLYWYSLKAYLAKTMAVDFQVCPVYVCSIVYWMTPGAPETSRLFL 505
 DB 453 FPOEKRMLIKERSGMYRLSSYFMARNVGDPLELALPTAFVFIYMWGLKPDPTTFIL 512

Db 428 FMAHVLGSLMELVLPASFLTFTYWMYVLRPGIVPPEFLTLLSVLLLYVLASQGLGALGAA 487
QY 528 SNSLOVATFVGPVTAIPVLLSFGFEVSKTIPITYLOWSSYLS-----YVRYGREG 577
Db 488 INDAKKASIVIVTMFLAVLTGYYVN--KVPFGWMMKYVSTTFYCYRLLVAIQSGSE 545
QY 578 VILTYGME-----RGDLTCLERCPCFPREPOSILRALDVEDAKLYMDFLVLFGLFFLALRL 633
Db 546 ETLRLMGCDCKKQAGSAATSAGCRFVEEVI-----GDVGMWTSVGVLFLEMFYRYL 599
QY 634 AYLVLYRYVK 643
Db 600 AYLAIR-RIK 608

RESULT 8

T46101
Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46101
R:Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Meves, H.W.; Lemcke, K.; Mayer, K.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23021
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-737 <ALC>
A:Cross-references: EMBL:AL132972
A:Experimental source: cultivar Columbia; BAC clone T25B15
C:Genetics:
A:Map position: 3
A:Introns: 122/1; 146/3; 225/2; 277/2; 338/3; 422/2; 535/1; 628/3; 664/3
A:Note: T25B15.80

Query Match 20.8%; Score 697; DB 2; Length 737;
Best Local Similarity 28.5%; Pred. No. 2.1e-46;
Matches 187; Conservative 126; Mismatches 250; Indels 92; Gaps 20;
QY 37 LKKVENHITAEQRFSLPKRSADVIEFVLSYV-REGPCWRKRGYKTLKLSGKFCRR 95
Db 122 LEDIEAATSVVKFOAEP--TFPIYLFIDITYKVTGK--MTSSSEKSLNGISGAYPG 178
QY 96 ELIGIMPAGCKSTPMNLAG--YRSGMKGOILVNGRP--RELTRFRKMSCYIMODDML 152
Db 179 ELLALMGPSGSKTLLNALGRFNQONIGGSVYNDKPYSKLKT---RIGFVTQDDVL 235
QY 153 LPHLTVLEAMVMSANLNL--TENPDVKNDLVTEILTALGLMSCSHTR-----TALLSGGO 205
Db 236 FPHLTVKETLTALTALLRLEKTLTEQEQRAASVIOELGLERCQDTMIGGSFVRVSGGE 295
QY 206 RRLALALELVNPPVPMFDEPTSGLDASCFQVVSMLKSLAOGGRTICTTHOPSAKLF 265
Db 296 RKRVCIGNEMTNPSLLLDDEPTSSLDSTALKIVQMLHCIAKAGKTIYVTHQPSRRLF 355
QY 266 EMFDKLYLSQGCIFKGVVYTNLPIKGLGHCPTYHNPADPIIEVASGEYGLDN-PML 324
Db 356 HREDKLVLSRGLLYFGKASEMSYFSSIGCSPLAMPAPFLLDLVNGNNDISVPSA 415
QY 325 FRA-----VQNGLC-----AMAEKKSPKEVNPAPCP-----PCPP 356
Db 416 LKEKMKIIRLEYVVRNVKCDVETQYLEEAYKTQIAVMKMKLMAVPVPLDEVKLMTCP- 474
QY 357 EVDPIESHFAVSTLTQFCILFKRTFELSLRDTVTLHFRMSSHVVGLVGLLYLHIGDD 416
Db 475 -----KRWGSLSWMEQYCLLSLRG--IKERRHDYFSLWRVTVQLSTAIILGLWMO--SDI 526
QY 417 ASKVFNNTGCLFSLMFLMFAALMPTVLTFFLEMAVEMREHLNWTSLKAYILAKTMADY 476
Db 527 TSQRPRTSGLLFIADVWGFPPVFTAIFFPQERAMLSKERESNMWRLSAYFVARTSD 586
QY 477 PFOVCPVVCYSIVYWMTGOPAEFSRFLFSALATATATAVAQSLGLLGAASNSLOVATF 536

RESULT 9

T21109
Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T21109
R:Thomas, K.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19375
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-695 <WIL>
A:Cross-references: EMBL:Z69635; NID:g1200023; PIDN:CAA93461.1; GSPDB:GN00022; CESP:F
A:Experimental source: clone F19B6
C:Genetics:
A:Gene: CESP:F19B6.4
A:Map position: 4
A:Introns: 9/1; 51/3; 167/2; 257/3; 515/3; 561/1
C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 20.5%; Score 686; DB 2; Length 695;
Best Local Similarity 27.5%; Pred. No. 1.4e-45;
Matches 184; Conservative 130; Mismatches 279; Indels 76; Gaps 17;

QY 31 PVL-----TTHLKKVENHITAEQRFSLPKRSADVIEFVLSYVSR----- 71
Db 45 PVLINDTPVASEIEQVEKELAHSSERFRPKPLSVTNDTESARIPSOVKTNAKKLVQNI 104
QY 72 EGPCKWRKRYK-TLLKCLSGKFCRRELIGIMPAGCKSTPMNLAGYRESGKM--GOIL 128
Db 105 QAVLLKKGVRQEIILKIDGVARPGLTETFMGSSGAGKTTLLNLTGRNLKNIETGDIM 164
QY 129 VNGRPRELTFRKMSCYIMODDMLPHLTVLEAMVMSANLNTENPDV-----KNDLVTE 183
Db 165 INGRNMISNMKKLSAYVOODDVFIGTLVRETFLRFAAKL---RSPSALGATELDSIVDE 221
QY 184 ILTALGLMSCSHTRTAL-----LSGGQRKRLAIALALELVNPPVPMFDEPTSGLDASCFQ 238
Db 222 LLVMSLKKCENTKVGTMTEKSLRGERKRLAFACELITDPPILFCDEPTSGLDSPMSHQ 281
QY 239 VVSLMKSLAOGGRTICTTHQPSAKLFEMFDPKLYLSQGCIFKGVVYTNLPIKGLGLH 298
Db 282 VIKALRQLTIEGTVICTTHQPSVSYVHMADQLLSQGHVAYAGPAKVDAFFRCGVP 341
QY 299 CPTYHNPADPIIEVASGE-----YGLNPMFLFRAVQNGLCAMAEKKS-----SPKNEVPAP 350
Db 342 IPKFVSSPDHFRMVRVISHKSFETEDDYNKRIEKTIVLEHDMKKEQSTHSTLSSRRRHET 401
QY 351 CPKPCPEVDPIESHFAVSTLTQFCILFKRTFELSLRDTVTLHFRMSSHVVGLVGLLY 410
Db 402 AP-----FTFPTWTWTAQFFIFORSSQLWRERSVLLVKLIQTLIIMSILGTY 450
QY 411 LHIGDDASKVFNNTGCLFSLMFLMFAALMPTVLTFFLEMAVEMREHLNWTSLKAYILA 470
Db 451 YGLEIDKSLPSFKGFAFVSQVMHMLFMMPAMTVFMKDYPVVVVREFQANMYSAYILA 510
QY 471 KTMADVPFOVCPVVCYSIVYWMTGOPAEF---SRFLFSALATATATAVAQSLGLLIGA 526
Db 511 KTTADSIQVLYVPFVIFSGILLGMTSLPYSVVIITNYLIINILLSNACSVGSGFAAMCGH 570

```

A:Residues: 1-725<STO>
A:cross-references: GB:AE002093; NID:q4262239; PIDN:AAD14532.1; GSPDB:GN
C:Genetics:
A:Gene: At2g01320
A:Map position: 2

Query Match      19.9%; Score 665.5; DB 2; Length 725;
Best Local Similarity 29.0%; Pred. No. 5.9e-44;
Matches 176; Conservative 120; Mismatches 250; Indels 61; Gaps:

Qy . . . 5 ALEANGCGL-----GPGAVAMVTLDEGAEPVLTTHLKKVNHHTEAQRFSHLPKPR
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 ALAAVAALLVRLFRAGP---IALLPEDEADDDYAET-----EDGGGDSIR-----

Qy 59 VDIETVEISYSVREGPCWKRKGYYTLAKCJSGKFCRRRELIGIMGPSGAGKSTFMTNLAA
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 VTIRNRNITCSLSDR---SSKSVRFLKNVSGEAKPGRLIAIMGPSGSGYTTLLINLVA
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 119 RESG---MKGOILVNGPRELRTFRKMSCYIMODDMLLPHLTVLEAMMYSAINLITE
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

125 LSLSPRLHLSGLLEVGKSPSSKAYK--LAPVROEDLFFSOLTIVRETLSFAAEQLQPLP
173 NPDVKNDLVTETLTALGLMSCSHT-----RTALLSGQKRRKRLAIALELVNNPVMFFEDD
183 SAEEDYVNNLLKGLVSCADSCVGDAKVGISGGKKRRLSLACELTASPSVIFADI
228 TSGLSASCFQVVSMLKSLAOGGRTIICTIHQPSAKLFEMFDKLYLTLSOGOCITFKGVV
243 TTGLDAFAQEKVMTLOKLQADGHTVICSIHQPRGSVYAKFDDIVLLTTEGLTVYAGPACAP
287 NLIPLYKLGJLHCPTYHNPADEIIEVASGEYGDLPMLF--RAVONGCLCMAEKKSS--
303 EPLTYFGNFGPCPEHVNPAEFLADLISVDYSSETSYSOKRKHVALVDAFORSRSSVI
342 -----PEKNEVPAPCPCCPEVDPIESHFTATSLTQFCILFKRTFLSILRLDVLV
363 ATPLSMKEETKNGM-----RPRKAIVERT--DGWNRQFFLLKRWANMQASRDGPTT
394 LREMSHVYVGLVILGLLYLHIGDDASKVFNNTGCLFFSNLFLMFAALMPTVLTFPLEMAA
414 VARMVSASAVIFGSVFMWKGKSTSIDRMRGLLQVAAINTAMAAUATKTTVGVFPEKRA
454 MREHUNYVYSKAYLAKTMADVPFOVVCPPVYCSIVYWMNTGQPAETSRFLFLFSALAT
474 DRERSKGSYGLPVLLSKRTIAEIDIGAAFLPMFGAVLYPMARLNPTLRSRFGKFCGIVTV
514 ALVAOSGLLIGCAASNSLOVATFVGPYTAIPVLLFSGFFVSFKTITYTLOWSSYLSYV
534 SFAASAGMLTVGAMVPSTEAMAVGPSLMTVFVIFGGTYVYNDNTPLIIFWIPRASLLI
574 GFEGVIL 580
594 AFOGLCI 600

```

RESULT 12

B96573
Protein F12M16.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: B96573
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
Nelson, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maity, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
Reference number: A86141; MUID:21016719; PMID:11130712

Accession: B96573

Status: preliminary

A:Molecule type: DNA

A:Residues: 1-590 <STO>

A:Cross-references: GB:AE005173; NID:g7769856; PIDN:AAF69534.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 19.8%; Score 663.5; DB 2; Length 590;
Best Local Similarity 31.7%; Pred. No. 6.5e-44;
Matches 105; Conservative 112; Mismatches 216; Indels 71; Gaps 18;

QY 82 KTLKCLSGKFCRRELIGIMPGSGACKSTFMNLAGYRESG-MKGQILVNGRPRELRTFR 140
DB 48 KVLKDVSCDASAEITAGPSGACKTLLLEILACKVSHGVSGQVNLVNGRPMDDPEYR 107
QY 141 KMSCYTQDDMLLPHLTVLEAMVSNLNL-TENPDVKNLDTVEITLALGLMSCSHRTA 199
DB 108 RVSGFPOEDALFPFVLTQETLTSYALLKTRKDAKAAK-VKRLIQELGLEHVDASRG 166
QY 200 L-----LSGQKRLAIALELVNPNVPMFDEPTSGDSCFQVYSLMKSIA-OGRTI 253
DB 167 QGSRGSGRRVRSIGVELVHDNPVILDEPTSGDSCFQVYSLMKSIA-OGRTI 226
QY 254 ICTIHPSAKLFEMFDFKILSGQCTFRGVNPNLIPYKLGCLHCPVTHNPADFTIEVA 313
DB 227 VLIHQPGFRILQIDRVLLSNGMVQVSGVYSLHOKIKFSGHQIPRRVNVLEYAIDIA 286
QY 314 SGEGDLNPNLFRVONGLCMAEK-KSPKNEVPAPCPPC-----PPEVDPIESHTFA 367
DB 287 -----GSLEPIRTQSCREISCYGHSKTWS-----CYISAGGELHQSOSH--S 327
QY 368 TSPLTQFCILFKRTFLSILRDTVLTHRMNVHVVIGVILGLYHLTGDD--ASKVFNNTG 425
DB 328 NSVLEEVQILGQSKNIFRTKOLFTRALQASIALGLISYLVNNGKKEAKVL-RTG 386
QY 426 CLFFSMLFMFAALMPTVLTPL--EMAVEMREHLNWNYSKAYVLAKTMADVPFQVVC 482
DB 387 FFAPILTLF-----LSSTEGPLFLQDRILMRETSRRAYRVLSYVLADTLFIPFLII 442
QY 483 PVVYCSIVVMTQPAETSRFLFSALATATLVAQSLGLLIGAASNSLQVATFVGPVTA 542
DB 443 SMLFATPVVWLVGLRRELDGFLYSLVIVLILMSNSF-----VACFSALVPNFINGTSV 497
QY 543 IPVL-----LFSGFFVSFKTIPTIYLOWSSYLSVVRGFEQVILTIYGMERGDTLCLEBC 597
DB 498 ISGIMSGFSLFGYFTAKRIPYWEFMHLSLFKYPFECLEKNEY---RGDV-----547
QY 598 PFREPOSILRALDVEDAKLYMDFLVIAGIFFLALRLAYLVLR 641
DB 548 -----FLKQDQKESQWSNLGIMASFTVGRVILGFFILWYR 584

RESULT 13

T08934

hypothetical protein F27G19.20 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000

C:Accession: T08934

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro

submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16519

A:Accession: T08934

A:Molecule type: DNA

A:Residues: 1-635 <BEV>

A:Cross-references: EMBL:AL078467; GSPDB:GN00062; ATSP:F27G19.20

A:Experimental source: cultivar Columbia; BAC clone F27G19

C:Genetics:

A:Gene: ATSP:F27G19.20

A:Map position: 4

A:Introns: 38/3; 253/1; 304/1; 414/3

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 19.1%; Score 640; DB 2; Length 635;
Best Local Similarity 27.5%; Pred. No. 4.9e-42;

Matches 186; Conservative 126; Mismatches 238; Indels 126; Gaps 25;

QY 29 EPPVLTH-----LKKVENHITFAQRFSLPKRSADVIEFVELSVSR-----EGPC 75
DB 11 ETPIAKTNDRLSLPFSIFKKNPVT-----LKFENLVTVTKLQDSQG-C 54

QY 76 WKRG---YKTLKCLSGKFCRRELIGIMPGSGACKSTFMNLAGYRESG---MKQILV 129
DB 55 FGKNDKTEERTILKGTIGTVKPGCEILAMLGSPSGKTSLLTALGVRGVGCKLGTNISY 114

QY 130 NGRPRELRTFRKMSCYIMODMLLPHLTVLEAMVSNLNLTENPDVKNLDTVE---ILT 186
DB 115 NKPPLS-KAVKRTTGEVTOQDALLYPNLTVTETLVFTALLRL-PNSFKKKEIKQAKAVT 172

QY 187 ALGLMSCSHRTA-----LLSGQKRLAIALELVNPNVPMFDEPTSGDSCFQVYVS 241
DB 173 ELGLDRCKDTIIGGPPFLRGVSGGERKRVSTIGOEILINPSLLFLDEPTSGDSTTAQRIVS 232

QY 242 LKMSLAOGGRTIICITHQPSAKLFEMFDFKILYLSQGCIFKGVVTVNLIPYKLGCLHCP 301
DB 233 ILWELARGRTVVTTHQP-----SKGNPVYFGLGSNAMDYFASVG-YSPL 277

QY 302 YH--NPADRTIEVASGEYGDLPNLMPLFRAYONGLCMAEKKSPEK-----344
DB 278 VERINPDSDELIDANK-----PLL--VISCPVSGVSDSQRPKAAALVAFYKTNLLD 330

QY 345 ----NEVPAPCPPC--PPEVDPIESHTFA--TSTLTQFCILFKRTFLSILRDTVLTHLRF 396
DB 331 SVINEVKGGDDLCNKPRESSRVATNTYGDWPTTWQFCVLLKRG-LKQRHDSFSGMKV 389

QY 397 MSHVVIGVILGLYHLIGDASKVFNNTGCLFFSMLFLMFAALMPTVLTFLPEMAVFMRE 456
DB 390 AQIFIVSFYLCGLLWMO--TKISRLQDQIGLLEFFISSEWAFPLFQOIFTFPQERAMLOKE 447

QY 457 HLNWYSLKAYVLAKTMADVPQVVCVIVVMTQPAETSRFLFSALATATLV 516
DB 448 RSGMTRFSPYFLSRVVGDPMLPILPTCTFLVITYMAGLNHNLANFFVTLVLVHVHVLV 507

QY 517 AOSGLLIGAASNSLQVATFVGPVTAIPVLLFSGFEVSEFKTIPTIYLOWSSYLSVYRGFE 576
DB 508 SGGGLGALCALYMDQKSAATLGLSVIMLTAGGYV--QHPVFTISWIKYISGYTYTK 565

QY 577 GVILTIYGM-----ERGDLYCLEERCPRFPQSILRALDVEDAKLYMDF-----LV 622
DB 566 LLILGOYTANELYPCGDNGKLR-----HVGDFEGIK-HITGNSGLVLSALA 610

QY 623 LGIFFLALRLAYLV 638

DB 611 LTAMLVVYRVIAIAL 626

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:46:34 ; Search time 49. seconds
(without alignments)
1361.086 Million cell updates/sec

Title: US-10-072-621-9

Perfect score: 3347

Sequence: 1 MAEKALAVCGGLPGAVAM.....FLALLLAVILRYVKSR 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*

2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*

7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*

10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*

12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*

13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3347	100.0	646	9 US-10-072-621-9	Sequence 9, Appli
2	3347	100.0	646	9 US-10-090-455-2	Sequence 2, Appli
3	3301	98.6	646	9 US-10-154-452-4	Sequence 4, Appli
4	3293	98.4	646	9 US-10-154-452-8	Sequence 8, Appli
5	3287	98.2	646	9 US-10-090-455-13	Sequence 13, Appli
6	3213	96.0	627	9 US-10-090-455-8	Sequence 8, Appli
7	2356	70.4	674	9 US-10-090-455-4	Sequence 4, Appli
8	2350	70.2	638	9 US-10-072-621-10	Sequence 10, Appli
9	879.5	26.3	663	9 US-10-108-605-245	Sequence 245, App
10	719	21.5	657	10 US-09-866-866A-14	Sequence 14, Appli
11	717	21.4	655	10 US-09-866-866A-27	Sequence 27, Appli
12	712	21.3	655	9 US-09-981-353-35	Sequence 35, Appli
13	712	21.3	655	9 US-10-120-687-61	Sequence 61, Appli
14	710	21.2	655	9 US-09-961-086-1	Sequence 1, Appli
15	704	21.0	655	9 US-10-090-455-5	Sequence 5, Appli
16	704	21.0	655	10 US-09-866-866A-10	Sequence 10, Appli
17	624.5	18.7	604	10 US-09-745-763-197	Sequence 197, App
18	585.5	17.5	651	9 US-10-090-455-6	Sequence 6, Appli
19	585.5	17.5	651	9 US-09-989-981A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1

US-10-072-621-9

; Sequence 9, Application US/10072621

; Patent No. US20020169137A1

; GENERAL INFORMATION:

; APPLICANT: Reiner, Peter B.

; APPLICANT: Connop, Bruce P.

; APPLICANT: Pollard, Michelle

; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION

; FILE REFERENCE: 100103.402

; CURRENT APPLICATION NUMBER: US/10/072.621

; CURRENT FILING DATE: 2002-02-08

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9

; TYPE: PRT

; LENGTH: 646

; ORGANISM: Homo sapiens

US-10-072-621-9

Query Match	100.0%;	Score	3347;	DB	9;	Length	646;
Best Local Similarity	100.0%;	Pred. No.	1.3e-293;				
Matches	646;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MAEKALAVCGGLPGAVAMAVTLEDGAEPPVLTTHLKKYVENHTEAQRSHLPKRSVD	60				
Db	1	MAEKALAVCGGLPGAVAMAVTLEDGAEPPVLTTHLKKYVENHTEAQRSHLPKRSVD	60				
Qy	61	IEFVELSVSYREGPCWRKRGYKTLKCLSKFCRRELIGIMGPSGAGKSTFMNLAGYRE	120				
Db	61	IEFVELSVSYREGPCWRKRGYKTLKCLSKFCRRELIGIMGPSGAGKSTFMNLAGYRE	120				
Qy	121	SGMKQILVNGRPRLRTFRKMSCYIMODMMLPLHVTLEAMVVSANLNTENPDVKNDL	180				
Db	121	SGMKQILVNGRPRLRTFRKMSCYIMODMMLPLHVTLEAMVVSANLNTENPDVKNDL	180				
Qy	181	VTELTALGLMSCSHTRTALLSGGQRRLATALELVNPNPVMFFDEPTSGLDASCFQV	240				
Db	181	VTELTALGLMSCSHTRTALLSGGQRRLATALELVNPNPVMFFDEPTSGLDASCFQV	240				
Qy	241	SLMKSQAQGGRTTICTITHQPSAKLFEMFDKLYILSGOCIFKGVVVTNLIPYLKGLGRCP	300				

Db 241 SLKSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVTLNIPYLKGLGLHCP 300
Qy 301 TYHNPADFIIEVASGEYGLDNLPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
Db 301 TYHNPADFIIEVASGEYGLDNLPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
Qy 361 IESHTATSTLTQFCILFKRTFSLILRDTVTLHREMSHVIVIGLIGLLYHIGDDASKV 420
Db 361 IESHTATSTLTQFCILFKRTFSLILRDTVTLHREMSHVIVIGLIGLLYHIGDDASKV 420
Qy 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Db 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Qy 481 VCPVVCISIVYWTGPAETSRELLFSALATATVAQSLGGLIGAASNSLQVATFVGPV 540
Db 481 VCPVVCISIVYWTGPAETSRELLFSALATATVAQSLGGLIGAASNSLQVATFVGPV 540
Qy 541 TAIPVLLFSGFFVSFKTIPTIYLQWSSYLSVRYGFEQVILTIYGMERDGLTCLLEERCPR 600
Db 541 TAIPVLLFSGFFVSFKTIPTIYLQWSSYLSVRYGFEQVILTIYGMERDGLTCLLEERCPR 600
Qy 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLRYRKSER 646
Db 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLRYRKSER 646

RESULT 2

US-10-090-455-2

; Sequence 2, Application US/10090455

; Publication No. US20030027259A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Hongyun

; APPLICANT: Le Bihan, Stephane

; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF

; FILE REFERENCE: 100103.406

; CURRENT APPLICATION NUMBER: US/10/090.455

; CURRENT FILING DATE: 2002-03-01

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 646

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-090-455-2

Query Match 100.0%; Score 3347; DB 9; Length 646;
Best Local Similarity 100.0%; Pred. No. 1.3e-293;
Matches 646; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 MAEKALEAVCGGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSADV 60
Db 1 MAEKALEAVCGGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSADV 60
Qy 61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMGPSGAGKSTFMNLAGYRE 120
Db 61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMGPSGAGKSTFMNLAGYRE 120
Qy 121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
Db 121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
Qy 181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSLGDSASCFQV 240
Db 181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSLGDSASCFQV 240
Qy 241 SLKSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVTLNIPYLKGLGLHCP 300
Db 241 SLKSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVTLNIPYLKGLGLHCP 300
Qy 301 TYHNPADFIIEVASGEYGLDNLPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360

Db 301 TYHNPADFIIEVASGEYGLDNLPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
Qy 361 IESHTATSTLTQFCILFKRTFSLILRDTVTLHREMSHVIVIGLIGLLYHIGDDASKV 420
Db 361 IESHTATSTLTQFCILFKRTFSLILRDTVTLHREMSHVIVIGLIGLLYHIGDDASKV 420
Qy 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Db 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Qy 481 VCPVVCISIVYWTGPAETSRELLFSALATATVAQSLGGLIGAASNSLQVATFVGPV 540
Db 481 VCPVVCISIVYWTGPAETSRELLFSALATATVAQSLGGLIGAASNSLQVATFVGPV 540
Qy 541 TAIPVLLFSGFFVSFKTIPTIYLQWSSYLSVRYGFEQVILTIYGMERDGLTCLLEERCPR 600
Db 541 TAIPVLLFSGFFVSFKTIPTIYLQWSSYLSVRYGFEQVILTIYGMERDGLTCLLEERCPR 600
Qy 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLRYRKSER 646
Db 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLRYRKSER 646

RESULT 3

US-10-154-452-4

; Sequence 4, Application US/10154452

; Publication No. US20020192821A1

; GENERAL INFORMATION:

; APPLICANT: Reiner, Peter B.

; APPLICANT: Roy, Josee

; APPLICANT: Connop, Bruce P.

; TITLE OF INVENTION: INCREASED FUNCTIONAL ACTIVITY AND/OR

; TITLE OF INVENTION: EXPRESSION OF ABC TRANSPORTERS PROTECTS AGAINST THE LOSS OF

; TITLE OF INVENTION: DOPAMINE NEURONS ASSOCIATED WITH PARKINSON'S DISEASE

; FILE REFERENCE: 100103.420

; CURRENT APPLICATION NUMBER: US/10/154.452

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 646

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-154-452-4

Query Match 98.6%; Score 3301; DB 9; Length 646;
Best Local Similarity 98.8%; Pred. No. 1.9e-289;
Matches 638; Conservative 3; Mismatches 5; Indels 0; Caps 0;

Qy 1 MAEKALEAVCGGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSADV 60
Db 1 MAEKALEAVCGGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSADV 60
Qy 61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMGPSGAGKSTFMNLAGYRE 120
Db 61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMGPSGAGKSTFMNLAGYRE 120
Qy 121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
Db 121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
Qy 181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSLGDSASCFQV 240
Db 181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSLGDSASCFQV 240
Qy 241 SLKSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVTLNIPYLKGLGLHCP 300
Db 241 SLKSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVTLNIPYLKGLGLHCP 300
Qy 301 TYHNPADFIIEVASGEYGLDNLPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
Db 301 TYHNPADFIIEVASGEYGLDNLPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360

QY 361 IESHTFATSTLTQRCILFKRTFLSILRDTVTLTHLRFMSHVHVIGVLGLLGLYHIGDDASKV 420
DB 361 IESHTFATSTLTQRCILFKRTFLSILRDTVTLTHLRFMSHVHVIGVLGLLGLYHIGDDASKV 420
QY 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
DB 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
QY 481 VCPVYCSIVYVMTGQPAETSRLFLSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
DB 481 VCPVYCSIVYVMTGQPAETSRLFLSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
QY 541 TAIPIVLLFSFGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
DB 541 TAIPIVLLFSFGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
QY 601 EPOQILRALDVEDAKLYMDFLVIGLFFLAURLLALVLRVYKSER 646
DB 601 EPOQILRALDVEDAKLYMDFLVIGLFFLAURLLALVLRVYKSER 646

RESULT 4
US-10-154-452-8
; Sequence 8, Application US/10154452
; Publication No. US20020192821A1
; GENERAL INFORMATION:
; APPLICANT: Reiner, Peter B.
; APPLICANT: Roy, Josee
; APPLICANT: Connop, Bruce P.
; TITLE OF INVENTION: INCREASED FUNCTIONAL ACTIVITY AND/OR
; TITLE OF INVENTION: EXPRESSION OF ABC TRANSPORTERS PROTECTS AGAINST THE LOSS OF
; TITLE OF INVENTION: DOPAMINE NEURONS ASSOCIATED WITH PARKINSON'S DISEASE
; FILE REFERENCE: 100103.420
; CURRENT APPLICATION NUMBER: US/10154.452
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-154-452-8

Query Match 98.4%; Score 3293; DB 9; Length 646;
Best Local Similarity 98.5%; Pred. No. 9.9e-289;
Matches 636; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
DB 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
QY 61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
DB 61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
QY 121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLEAMVYSANLNTENPDVKNL 180
DB 121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLEAMVYSANLNTENPDVKNL 180
QY 181 VTEILTALGLMSCSHTRTALLSGQQRKRLATALELVNPNPVFFDEPTSGLDSCFQV 240
DB 181 VTEILTALGLMSCSHTRTALLSGQQRKRLATALELVNPNPVFFDEPTSGLDSCFQV 240
QY 241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
DB 241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
QY 301 TYHNPADFIEVASGEYGLNPLMFRAVQNGLCMAEKSSPEKNEVPAPCPPEVDP 360
DB 301 TYHNPADFIEVASGEYGLNPLMFRAVQNGLCMAEKSSPEKNEVPAPCPPEVDP 360
QY 361 IESHTFATSTLTQRCILFKRTFLSILRDTVTLTHLRFMSHVHVIGVLGLLGLYHIGDDASKV 420
DB 361 IESHTFATSTLTQRCILFKRTFLSILRDTVTLTHLRFMSHVHVIGVLGLLGLYHIGDDASKV 420

DB 361 IESHTFATSTLTQRCILFKRTFLSILRDTVTLTHLRFMSHVHVIGVLGLLGLYHIGDDASKV 420
QY 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
DB 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
QY 481 VCPVYCSIVYVMTGQPAETSRLFLSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
DB 481 VCPVYCSIVYVMTGQPAETSRLFLSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
QY 541 TAIPIVLLFSFGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
DB 541 TAIPIVLLFSFGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
QY 601 EPOQILRALDVEDAKLYMDFLVIGLFFLAURLLALVLRVYKSER 646
DB 601 EPOQILRALDVEDAKLYMDFLVIGLFFLAURLLALVLRVYKSER 646

RESULT 5
US-10-090-455-13
; Sequence 13, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: 579, 598
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-090-455-13

Query Match 98.2%; Score 3287; DB 9; Length 646;
Best Local Similarity 98.5%; Pred. No. 3.5e-288;
Matches 636; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
DB 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
QY 61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
DB 61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
QY 121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLEAMVYSANLNTENPDVKNL 180
DB 121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLLPHLTVLEAMVYSANLNTENPDVKNL 180
QY 181 VTEILTALGLMSCSHTRTALLSGQQRKRLATALELVNPNPVFFDEPTSGLDSCFQV 240
DB 181 VTEILTALGLMSCSHTRTALLSGQQRKRLATALELVNPNPVFFDEPTSGLDSCFQV 240
QY 241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
DB 241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
QY 301 TYHNPADFIEVASGEYGLNPLMFRAVQNGLCMAEKSSPEKNEVPAPCPPEVDP 360
DB 301 TYHNPADFIEVASGEYGLNPLMFRAVQNGLCMAEKSSPEKNEVPAPCPPEVDP 360
QY 361 IESHTFATSTLTQRCILFKRTFLSILRDTVTLTHLRFMSHVHVIGVLGLLGLYHIGDDASKV 420
DB 361 IESHTFATSTLTQRCILFKRTFLSILRDTVTLTHLRFMSHVHVIGVLGLLGLYHIGDDASKV 420

Qy 421 FNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNWNYSKAYYLAKTMADVPFQV 480
Db 421 FNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNWNYSKAYYLAKTMADVPFQV 480
Qy 481 VCPVVCYSIVYVMTGQPAETSRFLFSALATATATVAQSLGLLIGAASNSLQVATFYGPV 540
Db 481 VCPVVCYSIVYVMTGQPAETSRFLFSALATATATVAQSLGLLIGAASNSLQVATFYGPV 540
Qy 541 TAIPVLLFSGFFVSFKTIPYIYLQWSSYLSYVRGFEVGVILTIYGMERGDLTCLLEERCXFR 600
Db 541 TAIPVLLFSGFFVSFKTIPYIYLQWSSYLSYVRGFEVGVILTIYGMERGDLTCLLEERCXFR 600
Qy 601 EPOSILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSR 646
Db 601 EPOSILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSR 646

RESULT 6

US-10-090-455-8
: Sequence 8, Application US/10090455
: Publication No. US20030027259A1
: GENERAL INFORMATION:
: APPLICANT: Chen, Hongyun
: APPLICANT: Le Bihan, Stephane
: TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
: FILE REFERENCE: 100103.406
: CURRENT APPLICATION NUMBER: US/10/090.455
: CURRENT FILING DATE: 2002-03-01
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 627
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-090-455-8

Query Match 96.08; Score 3213; DB 9; Length 627;
Best Local Similarity 98.98; Pred. No. 1.6e-281;
Matches 620; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Qy 20 MAVTLEDGAPPVLTTHLKKVENHITAEQRFSLPKRSADVIEFVELSYSVREGPCWRKR 79
Db 1 MAVTLEDGAPPVLTTHLKKVENHITAEQRFSLPKRSADVIEFVELSYSVREGPCWRKR 60
Qy 80 GYKTLKCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTF 139
Db 61 GYKTLKCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTF 120
140 RKMSCIYIMQDDMLLPHLTVLLEAMMVSANLNTENPDVKNLDVTEILTALGLMCSHTRTA 199
121 RKMSCIYIMQDDMLLPHLTVLLEAMMVSANLNTENPDVKNLDVTEILTALGLMCSHTRTA 180
Qy 200 LLSGGQKRRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHQ 259
Db 181 LLSGGQKRRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHQ 240
Qy 260 PSAKLFEMFDKLYTLISOGQCIKFGVVTNLPYLKGLGHCPTVHNPADEFIEVASGEYGD 319
Db 241 PSAKLFEMFDKLYTLISOGQCIKFGVVTNLPYLKGLGHCPTVHNPADEFIEVASGEYGD 300
Qy 320 LNPMLFRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDPVIESHTATSLTQFCILFK 379
Db 301 LNPMLFRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDPVIESHTATSLTQFCILFK 360
Qy 380 RTFLSLIRDRVLRHLRFSMSHVIVGLVGLYHIGDDASQVFNNTGCLFTSMLFLMFAAL 439
Db 361 RTFLSLIRDRVLRHLRFSMSHVIVGLVGLYHIGDDASQVFNNTGCLFTSMLFLMFAAL 420
Qy 440 MPTVLTLEMAVFMREHLNWNYSKAYYLAKTMADVPFQVCPVVCYSIVYVMTGQPAE 499
Db 421 MPTVLTLEMAVFMREHLNWNYSKAYYLAKTMADVPFQVCPVVCYSIVYVMTGQPAE 480

Qy 500 TSRELLFSALATATVAQSLGLLIGAASNSLQVATFVGPVTAIPVLLFSGFFVSFKTIP 559
Db 481 TSRELLFSALATATVAQSLGLLIGAASNSLQVATFVGPVTAIPVLLFSGFFVSFKTIP 540
Qy 560 TYLQWSSYLSYVRGFEVGVILTIYGMERGDLTCLLEERCPPREPOSILRALDVEDAKLYMD 619
Db 541 TYLQWSSYLSYVRGFEVGVILTIYGMERGDLTCLLEERCPPREPOSILRALDVEDAKLYMD 600
Qy 620 FLVLGIFFLALRLAYLVLRYVKSR 646
Db 601 FLVLGIFFLALRLAYLVLRYVKSR 627

RESULT 7

US-10-090-455-4
: Sequence 4, Application US/10090455
: Publication No. US20030027259A1
: GENERAL INFORMATION:
: APPLICANT: Chen, Hongyun
: APPLICANT: Le Bihan, Stephane
: TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
: FILE REFERENCE: 100103.406
: CURRENT APPLICATION NUMBER: US/10/090.455
: CURRENT FILING DATE: 2002-03-01
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 674
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-090-455-4

Query Match 70.48; Score 2356; DB 9; Length 674;
Best Local Similarity 70.18; Pred. No. 4.6e-204;
Matches 454; Conservative 80; Mismatches 76; Indels 38; Gaps 4;
Qy 26 DGAEPVLTTHLKKVENHITAEQRFSLPKRSADVIEFVELSYSVREGPCWRKRKYKTL 85
Db 38 ENTETDLNGLHLKVDNLTAEQRFSSLPRAAVNIEFRDLSYVPEGPWRRKGYKTL 97
Qy 86 KCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTFRKMSY 145
Db 98 KGISGKFNSELVAIMGSPGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTFRKMSY 157
Qy 146 IMQDDMLLPHLTVLLEAMMVSANLNTENPDVKNLDVTEILTALGLMCSHTRTALLSGQ 205
Db 158 IMQDDMLLPHLTVOEAMMVSANLNTENPDVKNLDVTEILTALGLMCSHTRTALLSGQ 217
Qy 206 RKRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHOPSALF 265
Db 218 RKRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHOPSALF 277
Qy 266 EMFDKLYLSOGQCIKFGVVTNLPYLKGLGHCPTVHNPADEFIEVASGEYGDNLPMFL 325
Db 278 ELFDOLYLSOGQCIKFGVVTNLPYLKGLGHCPTVHNPADEFIEVASGEYGDNLPMFL 337
Qy 326 RAVQNGLCAMAEKSSPEKNEVPAPCPPEVDPVIESHTATSLTQFCILFK 361
Db 338 RAVREGMC-----DSHKKRDLGGDA-----EVNPFLLWHRPSEVKOTKRKLGRKDKSSS 386
Qy 362 --ESHTATSLTQFCILFKRTFLSLIRDRVLRHLRFSMSHVIVGLVGLYHIGDDASK 419
Db 387 MEGCHSFSASCLTQFCILFKRTFLSLIRDRVLRHLRFSMSHVIVGLVGLYHIGDDASK 446
Qy 420 VFNNTGCLFTSMLFLMFAALMPTVLTFFLEMAVFMREHLNWNYSKAYYLAKTMADVPFQ 479
Db 447 VLSNSGFFFSMLFLMFAALMPTVLTFFLEMAVFMREHLNWNYSKAYYLAKTMADVPFQ 506
Qy 480 VVCPVVCYSIVYVMTGQPAETSRFLFSALATATVAQSLGLLIGAASNSLQVATFYGP 539
Db 507 IMFVVCYSIVYVMTGQPAETSRFLFSALATATVAQSLGLLIGAASNSLQVATFYGP 566
Qy 540 VTAIPVLLFSGFFVSFKTIPYIYLQWSSYLSYVRGFEVGVILTIYGMERGDLTCLLEERC 598

Db 567 VTAIPVLLSGFVSFDTIPTTYLQWMSYISYVRGEGVILSYGLDREDLHCDIDETCH 626
Qy 599 FREPOSILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSER 646
Db 627 FORSEAILRELDVENAKLYLDFVLGIFFLISURLIAYLVLRYKIRAER 674

RESULT 8
US-10-072-621-10
; Sequence 10, Application US/10072621
; Patent No. US20020169137A1
; GENERAL INFORMATION:
; APPLICANT: Reiner, Peter B.
; APPLICANT: Connop, Bruce P.
; APPLICANT: Pollard, Michelle
; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION
; FILE OF INVENTION: BY MODIFICATION OF ABC TRANSPORTER EXPRESSION OR ACTIVITY
CURRENT APPLICATION NUMBER: US/10/072,621
; CURRENT FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-621-10

Query Match 70.2%; Score 2350; DB 9; Length 638;
Best Local Similarity 69.9%; Pred. No. 1.5e-203;
Matches 453; Conservative 80; Mismatches 77; Indels 38; Gaps 4;

Qy 26 DGAEPPVLTHLKKVNHTEAQRSHLPKRSAVDIEFVELSYSVREGPCWRKRGYKTL 85
Db 2 EATETOLLNGLHKKVDNNTTEAQRSSLPRAAVNIEFRDLSYSPGEGWRRKGYKTL 61
Qy 86 KCLSGKFCRRELIGIMPSCGACKSTFMNLAGYRESGMRGQILVNGRPRELRTFRKMSY 145
Db 62 KGISGRFNSGELVAIMPSCGACKSTLMNLAGYRETMKGAVLINGLPRDLRCFRKVS 121
Qy 146 IMQDDMLLPHLTVLEAMVYSANLNTENPDVKNLDTTEILTAGLMSCHSTRFALLSGG 205
Db 122 IMQDDMLLPHLTVLEAMVYSAHLKQEKDEGREMVKETLTAGLMSCHSTRFALLSGG 181
Qy 206 RKRLATALELVNPPVPMFTDEPTSGDSDSCFOVSLMKSLAOGRTICTIHOPSAKLF 265
Db 182 RKRLATALELVNPPVPMFTDEPTSGDSDSCFOVSLMKSLAOGRTICTIHOPSAKLF 241
Qy 266 EMFDKLYLSQGCIFKGVVTVNLIPYLKGLGHCPTTHNPADPIIEVASGEYGLNPMFL 325
Db 242 ELFDQLYLSQGCYVYRGKVCNLVPLRLDLGLNCPYHNPADPFVMEVASGEYGDQNSRLV 301
Qy 326 RAVQNGLCMAEKKSSPERNEVPAPCPPEVDPT----- 361
Db 302 RAVREGMC-----DSHDKRDLDGDA-----EYNPLFHLRSPBEVKOTRLKGLRDKDSS 350
Qy 362 --ESHFTFANSTLTFQFCILKRTPLSLTRDTVLTHLRFMSHVIGVLGLYHIGDDASK 419
Db 351 MECCHSFSASCTQFCILKRTPLSLTRDSTLMDRSLVTLHRTSHIGIGLIGLIGIGNEAK 410
Qy 420 VFNNCTGLFMSLFLMFAALMPTVLTFFLEMAVMEHRLNLYWYSLKAYYLAKTMAVDFPQ 479
Db 411 VLNSGFLFMSLFLMFAALMPTVLTFFLEMGVFLREHLNLYWYSLKAYYLAKTMAVDFPQ 470
Qy 480 VVCPVVCSTVYVMTQPAETSRFLFSALATATAYLAQSLGLLGAANSLSQVATFVGP 539
Db 471 IMPFVAYCSTVYVMTSQPSDAVAFVLAALGTMTSLVAQSLGLLGAANSLSQVATFVGP 530
Qy 540 VTAIPVLLSGFVSFDTIPTTYLQWMSYISYVRGEGVILSYGLDREDLHCDIDETCH 598
Db 531 VTAIPVLLSGFVSFDTIPTTYLQWMSYISYVRGEGVILSYGLDREDLHCDIDETCH 590

Qy 599 FREPOSILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSER 646
Db 591 FQKSEAILRELDVENAKLYLDFVLGIFFLISURLIAYLVLRYKIRAER 638

RESULT 9
US-10-108-605-245
; Sequence 245, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCOD
; FILE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 245
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-108-605-245

Query Match 26.3%; Score 879.5; DB 9; Length 663;
Best Local Similarity 33.9%; Pred. No. 1.1e-70;
Matches 206; Conservative 121; Mismatches 237; Indels 43; Gaps 12;

Qy 73 GPCWRK-----RGY-----KTLKCLSGKFCRRELIGIMPSCGACKSTFMNLA 116
Db 61 GSGWRQLVNRTRGLFCNERHIPAPRKHLLKNVGVAYPGELLAVMGSSGAGKTTLLNALA 120
Qy 117 GYRESGMRK-----QGIILVNGRPRELRTFRKMSYIMQDDMLLPHLTVLEAMVYSANLNT 172
Db 121 FRSPQIGVSPSGMRLLNGOPVDAKEMQARCAVQDDDLFGSLTAHEHLIFQAMVMPR 180
Qy 173 NPQVKNDL--VTBILTAGLMSCHT-----RTALLSGGQRKRLAIALBELVNNPVMFF 224
Db 181 HLTQRVARVDQVQIELSLKCOHTITIGVGRVKGLSGGERKRLAFASALTDPPLLIC 240
Qy 225 DEPTSGDSDSCFOVSLMKSLAOGRTICTIHOPSAKLFEMFDKLYLSQGCIFKGV 284
Db 241 DEPTSGDSDSTAHSVVQVLKLSQKGTVLTITIHOPSSSELFELDKILLMAEGRVAPLGT 300
Qy 285 VTNLIPYLKGLGHCPTTHNPADPIIEVAS--GEYGLNPLMFRAVQNGLCMAEKKSS 341
Db 301 PSEAVOFFSVGAQCPTNYPADFYVOVLAVPGRITESRDRIAKICDN----FAISKVA 356
Qy 342 PEKNEVPAPCPPEVDPTIESHTFATSTLTQFCILKRTPLSLTRDTVLTHLRFMSHV 401
Db 357 RDMQELLATKNLEKPLEQEPNGYKATWFMQFRAVLWRSWLSVLKPLLVKRLIQTMM 416
Qy 402 IGVLLGLLYHIGDDASK--VFNNCTGLFMSLFLMFAALMPTVLTFFLEMAVMEHRLN 459
Db 417 VAILIGLIEL--GOQLTVGVGMNINGAIFLNTMTFQNVFATINVTSELPVFMREARS 474
Qy 460 YWYSLRAYYLAKTMAVDFPQVVCVYVYVMTQPAETSRFLFSALATATAYLAQ 519
Db 475 RLTRCDTYFGLGTIAELPLFPLVPLTAIYPMIGLRAGVLHFFNCLAVLTVANVSTS 534
Qy 520 LGLLGAANSLSQVATFVGPVTAIPVLLSGFVSFDTIPTTYLQWMSYISYVRGEGV 579
Db 535 FGVLLSCASSSTSMALSGVPPVILPFLFGFFLNSGVPVYLKWLKLSYLSWFRYANEGLL 594
Qy 580 LTIYG-MERGDLCLEER--CPFREPOSILRALDVEDAKLYMDLVLGIFFLALRLAYL 636

Db 595 INOWADVERGEISCTSSNTTCP--SSGKVIETLNFSAADPLDYVGLAILVSRVLAYL 653
QY 637 VLRYRVK 643
Db 654 ALRLRAR 660

RESULT 10

US-09-866-866A-14
; Sequence 14, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 14
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-866-866A-14

Query Match 21.5%; Score 719; DB 10; Length 657;
Best Local Similarity 29.1%; Pred. No. 3.3e-56;
Matches 183; Conservative 133; Mismatches 258; Indels 54; Gaps 16;

QY 61 IEFVELSY--SVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGY 118
Db 37 LSFHHITYRVKVGSLVRKTEKEILSDING-IMKPGNLAILGPTGGKSSLLDVLAR 95
QY 119 RE-SGKMGQILVNGRPRLRTFRKMSCYIMQDDMLPLHLTVLEAMVVSANLN--TENPD 175
Db 96 KDPKGLSGDVLNGAPQPAH-FKCCSGYVQDDVVMGTLTVRENLOFSAALRLPTMKNH 154
QY 176 VKNDLVTETLTALGLMSCSHTRTAL-----LSGQGRKRLAIALEILVNNPVMFDEPTSG 230
Db 155 ENERNITIKELGLEKADVSKVGTFIRGISGGERKRTSIGMELTIDPSILFLDEPTG 214
QY 231 LDSASCFQVYSLMKSLAOGGRTIICTHOPSAKLFEMFDKLYILSQGOCIFKGVVNTLP 290
Db 215 LDSSTANAVLLLLKRMKSQGRITIFSIHQPRYSIFKLFDSLLASGKLVHGPQAQALE 274
QY 291 YLKGGLGHCPYHNPAADFIEVASGEYGDNLPMFLFRAVQNGLCAMAEKKSPEKNEVP-- 348
Db 275 YFASAGYHCEPYNNAADFLLVINGDSSAV--MLAREQDN---EANKTEEPSKGEKPI 329
QY 349 -----APCPPCPPVEDPI-----ESHFTATSLTQFCILFKRTFLSTL 386
Db 330 ENLSSEYINSALYGETKALDQLPGAQEKKGTSARKEPVYVTSFCHQLRWIARRSPKNLL 389
QY 387 RDTVLTHLRFMSHVIGVLGLLYLHIGDDASKVFNNMTGCLFFSMLFLMFAALMPTVLT 446
Db 390 GNPOASVAQLIVTVILGLIIGAIYFDLKYDAGMONRAGVLFFLTNNOCFSSVSAVEL-F 448
QY 447 PLEMAVFMREHLNRYWSLAKYLAKTMAAD-VPFQVVCVVYCSIVVWMTGQPAETSRFL 505
Db 449 VVEKKLFIHEYISGYRVSSYFFGKVMSDLPMRELPSPVIFTCIYFMLGLKKTVDAPFI 508
QY 506 FSALATATVAOSGLGLLIGAASNSLOVATFPGVTAIPVLLFSGFFVSKTIPYLOW 565
Db 509 MMFTLIWAYTASSALATATQSVSVATLMTWTAIFVFMFLFSGLLVWERTIGPWSLW 568
QY 566 SYLSYRVYGFEGVILTIY-GME-----RGDLTCLEER--CPFREPQSIILRALDVEDAK 615

Db 569 QYFSIPRYGFTALQYNEFLGQEPFCPGFNVTDNSTCVNSYAICTGNE-YLINOQIELSPWG 627
QY 616 LYMDFLVLGIFFLALRLALLAYLVLRVVK 643
Db 628 LMKNHVALACMIIFLTIAIYKLLFLKK 655

RESULT 11

US-09-866-866A-27
; Sequence 27, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 27
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-27

Query Match 21.4%; Score 717; DB 10; Length 655;
Best Local Similarity 28.0%; Pred. No. 4.9e-56;
Matches 178; Conservative 141; Mismatches 245; Indels 72; Gaps 16;

QY 61 IEFVELSYSV--REG--PCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLA 116
Db 37 LSFHHITYRVKVGSLGPLPC-RAPVEKEILSNING-IMKPGNLAILGPTGGKSSLLDVL 94
QY 117 GYRE-SGKMGQILVNGRPRLRTFRKMSCYIMQDDMLPLHLTVLEAMVVSANLN--TEN 173
Db 95 ARKDPESGLDVLNGAPRP-ANFKCNSGVVQDDVVMGTLTVRENLOFSAALRLATWT 153
QY 174 PDVKNDLVTETLTALGLMSCSHTRTAL-----LSGQGRKRLAIALELVNPPVMFDEPT 228
Db 154 NHEKNERINRVIOELGLDKVADSKVGTFIRGVSGGERKRTSIGMELTIDPSILFLDEPT 213
QY 229 SGLDASCFQVYSLMKSLAOGGRTIICTHOPSAKLFEMFDKLYILSQGOCIFKGVVNTL 288
Db 214 TGLDSTANAVLLLLKRMKSQGRITIFSIHQPRYSIFKLFDSLLASGKLVHGPQAQ 273
QY 289 IPLYKGLGHCPYHNPAADFIEVASGEYGDNLPMFLFRAVQNGLCAMAEKKSPEKNEVP 348
Db 274 LOYFESAGYHCEAYNNPAADFLLDINGO-----STAVALNREEDFKATEII 319
QY 349 ACPPCPPVEDPI-----ESH-----TFATSLTQFCI 376
Db 320 EPSKQDKPLIEKLAETIYVNSSFYKETKAEHLQSGGKKKKTIVFKETISYTSFCHQLRW 379
QY 377 LFKRFTSLTRDTVLTHLRFMSHVIGVLGLLYLHIGDDASKVFNNMTGCLFFSMLFLMF 436
Db 380 VSKRFSKNNLPQASIAOIIYVTVGLGLVIGAIYFGLKNDSTGIONRAGVLEFLTTNQCF 439
QY 437 AALMPTVLTVPLEMAVFMREHLNRYWSLAKYLAKTMAAD-VPFQVVCVVYCSIVVWMTG 495
Db 440 SSVSAVEL-FVVEKKLFIHEYISGYRVSSYFGLGKLLSLLPMRLPSIIFTCIYVFM 498
QY 496 QPAETSRFLFLSALATATATVAOSGLGLLIGAASNSLOVATFPGVTAIPVLLFSGFFVS 555
Db 499 LKAKADAFVFMFTLMWAVSASSMALATAAGQSVSVATLMTTCFVFMFIIFSGLLVNL 558

Qy 556 KTIPTYLQWSYLSYVRGREGVLTIIY-----GME-RGDLTCLERBCRFEPQSIILR 607
Db 559 TTASLWSLQYFSIPRYGTALQHNBEFLGONFCPLGNATNCNAYATCT-GEELYVKQ 617
Qy 608 ALDVEDAKLYMDFVLVLGIFFLALRLLAYLVLYRVRK 643
Db 618 GIDLSPGLWKNHVALACMIVIFUTIIAYLKLFLKK 653

RESULT 12
US-09-981-353-35
; Sequence 35, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 35
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
US-09-981-353-35

	Best Match	21.3%	Score 712;	DB 9;	Length 655;	
	Best Local Similarity	27.8%	Prod No 1.4e-55;			
	Matches 177; Conservative	141;	Mismatches 246;	Indels 72;	Gaps 16;	
Qy	61 IEFVELSYV--REG--PCWRKRGYTLKLCLGSCFCRRELIGITMGSPGAGKSTFMILA	116				
		:::::~::~:				
Dd	37 LSFHNICYRVKLKSGGLPCC-RKPVEKEILSNING-IMKPGLNATLGTGGCKSSLLOVLA	94				

```

Db      559  TTIASWLSNLOQYESIPRYGTALQHNEFLQGNPCGLNATGNPNCAVTCT -GEEYLVKQ  617
Qy      608  ALDVEDAKYMDFLVLGICIFFFLALRLLAYLVLRYVK  643
      :|: : : | : :|:| | : |
Db      618  GIDLSPWGLMKNHVALACMIVIFLTIAYLKLLFLKK  653

RESULT 13
US-10-120-687-61
; Sequence 61, Application US/10120687
; Publication No. US20030082155A1
; GENERAL INFORMATION:
; APPLICANT: Massachusetts General Hospital
; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treati
; TITLE OF INVENTION: Mellitus
; FILE REFERENCE: 3284/1235B
; CURRENT APPLICATION NUMBER: US/10/120,687
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: US60/169082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 09/963,875
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/215109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/238880
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/731261
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-120-687-61

```

[illegible]

Db 499 LKPKADAFFVMMFTLMVAVSASSMALAIAAGQSVSVATLLMTICFVFMIFSGLLVNL 558
Qy 556 KTIPTYLOWSSYLSYRYGEGVILTY-----GME-RGDLTCLERCPFRPQSILR 607
Db 559 TTIASWLSWLQYFSIPRYGFTALQHNFLQNCFCPLNATGNNPCNYATCT-GEEYLVKQ 617
Qy 608 ALDVEDAKLYMDFLVLGIFFLALRLAYLRLRYRVK 643
Db 618 GIDLSPWGLKKNHVALACMIVFIETIAYLKLLFLKK 653

RESULT 14.

US-09-961-086-1
; Sequence 1, Application US/09961086
; Publication No. US20030036645A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynde
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; FILE REFERENCE: WHICH ENCODES IT
; CURRENT APPLICATION NUMBER: US/09/961.086
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1

Query Match 21.2%; Score 710; DB 9; Length 655;
Best Local Similarity 27.8%; Pred. No. 2.1e-55;
Matches 177; Conservative 140; Mismatches 247; Indels 72; Gaps 16;
Qy 61 IEFVELSV--REG--PCWRKRGYKTLKCLSGKFCRRELIGMPSGAGKSTFMNLA 116
Db 37 LSFHNICRYVKLSGFLPC-RKPEKEILSNING-IMKPGNALIGPTGGKSSLLDVA 94
Qy 117 GYRE-SGMKGQILVNGRPRELRTFRKMSCYIMODMLLPHLTVLEAMVMSANL--TEN 173
Db 95 ARKDPSGLSGDLVINGAPR-ANFKCNSGVYVQDDVVMGTLTVRENLFQESAALRLATTMT 153
Qy 174 PDVKNLVTILALGLMSCSHRTAL-----LSGGORKRLATALELVNPPVFFDEPT 228
Db 154 NHEKNERINRVIELGLDKVADSKVGQFTIRGVSGGERKRTSIGMELITDPSILFDEPT 213
Qy 229 SGDSASCFVWSLMSLAOGGRTIICTIHQPSAKLFEMEDKLYLSOGOCIFKGVVNTL 288
Db 214 TGLDSSANAVALLKRLKMSKQRTIIFSIHQPRYSIFKFLDLSLTLASGRMLFHGPAQEA 273
Qy 289 IPYKGLGLHCPYHNPADFIIEVASGEYGDLPNMFRAVONGLCMAEKKSSPEKNEVP 348
Db 274 LGVFESAGYHCEAYNPNADFLDING-----STAVALNREEDFKATEII 319
Qy 349 APCPPCPPEVDPI-----ESH-----TATSTLTQFCI 376
Db 320 EPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKKITVFKESITVTSFCHOLRW 379
Qy 377 LFKRTFLSILRDVTLTHLRFMSHVIGVLGLYHLHIGDDASKVFNNTGCLFFSMLFLMF 436
Db 380 VSKRSFNKLLGNQASQAIIIVTVVLGLVIGAIYFGLKNDSTGIONRAGVLFELTTNQCF 439
Qy 437 AALMPTVLTLPLEMAVFMREHLNWSLYSKAYYLAKTMAV-VPEQVCPVVCISIVYMTG 495
Db 440 SSVSAVEL-FVVEKKLFIHEYISGYRVSSYFLGKLLSLLDPMRLPSIIFTICIVYEMLG 498

Qy 496 QPAETSRFLFSALATATATAVAOSLGLLIGAANSLOVATFVGPTAIPVLLFSGFFVSF 555
Db 499 LKPKADAFFVMMFTLMVAVSASSMALAIAAGQSVSVATLLMTICFVFMIFSGLLVNL 558
Qy 556 KTIPTYLOWSSYLSYRYGEGVILTY-----GME-RGDLTCLERCPFRPQSILR 607
Db 559 TTIASWLSWLQYFSIPRYGFTALQHNFLQNCFCPLNATGNNPCNYATCT-GEEYLVKQ 617
Qy 608 ALDVEDAKLYMDFLVLGIFFLALRLAYLRLRYRVK 643
Db 618 GIDLSPWGLKKNHVALACMIVFIETIAYLKLLFLKK 653

RESULT 15

US-10-090-455-5
; Sequence 5, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-5

Query Match 21.0%; Score 704; DB 9; Length 655;
Best Local Similarity 27.7%; Pred. No. 7.4e-55;
Matches 176; Conservative 141; Mismatches 247; Indels 72; Gaps 16;

Qy 61 IEFVELSV--REG--PCWRKRGYKTLKCLSGKFCRRELIGMPSGAGKSTFMNLA 116
Db 37 LSFHNICRYVKLSGFLPC-RKPEKEILSNING-IMKPGNALIGPTGGKSSLLDVA 94
Qy 117 GYRE-SGMKGQILVNGRPRELRTFRKMSCYIMODMLLPHLTVLEAMVMSANL--TEN 173
Db 95 ARKDPSGLSGDLVINGAPR-ANFKCNSGVYVQDDVVMGTLTVRENLFQESAALRLATTMT 153
Qy 174 PDVKNLVTILALGLMSCSHRTAL-----LSGGORKRLATALELVNPPVFFDEPT 228
Db 154 NHEKNERINRVIELGLDKVADSKVGQFTIRGVSGGERKRTSIGMELITDPSILSDEPT 213
Qy 229 SGDSASCFVWSLMSLAOGGRTIICTIHQPSAKLFEMEDKLYLSOGOCIFKGVVNTL 288
Db 214 TGLDSSANAVALLKRLKMSKQRTIIFSIHQPRYSIFKFLDLSLTLASGRMLFHGPAQEA 273
Qy 289 IPYKGLGLHCPYHNPADFIIEVASGEYGDLPNMFRAVONGLCMAEKKSSPEKNEVP 348
Db 274 LGVFESAGYHCEAYNPNADFLDING-----STAVALNREEDFKATEII 319
Qy 349 APCPPCPPEVDPI-----ESH-----TATSTLTQFCI 376
Db 320 EPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKKITVFKESITVTSFCHOLRW 379
Qy 377 LFKRTFLSILRDVTLTHLRFMSHVIGVLGLYHLHIGDDASKVFNNTGCLFFSMLFLMF 436
Db 380 VSKRSFNKLLGNQASQAIIIVTVVLGLVIGAIYFGLKNDSTGIONRAGVLFELTTNQCF 439
Qy 437 AALMPTVLTLPLEMAVFMREHLNWSLYSKAYYLAKTMAV-VPEQVCPVVCISIVYMTG 495
Db 440 SSVSAVEL-FVVEKKLFIHEYISGYRVSSYFLGKLLSLLDPMRLPSIIFTICIVYEMLG 498
Qy 496 QPAETSRFLFSALATATATAVAOSLGLLIGAANSLOVATFVGPTAIPVLLFSGFFVSF 555
Db 499 LKPKADAFFVMMFTLMVAVSASSMALAIAAGQSVSVATLLMTICFVFMIFSGLLVNL 558

QY	556	KTIPTYLQWSSYLYRYGEGVILTY-----GME-RGDLTCLERCPFREPOSILR	607
Db	559	TTIASWLSWLOYSIPRYGFTALQHNFLGQNFPCGLNATGNPCNYATCT-GEEYLVKQ	617
QY	608	ALDVEDAKLYMDFLVLGIFFLALRLAYLVLYRVK	643
Db	618	GIDLSPWGLKKNHVALACMIVIFLTAYKLLFLKK	653

Search completed: June 10, 2003, 16:53:11
Job time : 51 secs

Query Match	32.3%	Score 1079.5	DB 5	Length 576
Best Local Similarity	41.6%	Pred. No. 7.9e-72		
Matches	226	Conservative 84	Mismatches 132	Indels 101
Gaps	6			
QY	46	EAQRFSLPKRSADVIEFVLSYVRGEGCWRKRGYKTLKLCLSGKFCRRELIGIM	PGS	105
DB	31	QPKTLQHLKPRPAVDLAFHNLTVRKEG---	NRSNAKTILKGVSGRLSRGELTA	IMPGS 87
QY	106	AGKSTPMNLIAQYREGSMKGOIILVNRPRELTFTRKMSCYIMQDDMLPPLHVL	EAMMVS	165
DB	88	AGKSTLLNLSGYKTSIEGVSMTMGAERNLAFRKLISAYIMQDNQHLNLT	VOEAMTVA	147
QY	166	ANLNLTET---NPQVKNDLVTEILTALGLMSCSHRTALLSGGOKRLAI	ALELVNPNPVM	222
DB	148	TNLKLSKKFSKPE-KHSMIDDILLTSLSEHRYTMTNRNLGGOKKRLSIA	LELVSNPIM	206
QY	223	FFDEPTSGLDSASCFQVVSUMKSLAOGGRTIICTIHQPSAKLFEMFDKLYL	SGQCQIFK	282
DB	207	FFDEPTSGLDSSTCFQCQIHLKLLAAGRTVICTIHQPSARLFEFMDQVLT	ADGQCQVYQ	266
QY	283	GVVYNLIPYLKGLGLHCPPTHYNPADFIIEVASGEYGDLPNMLPRAV---		328
DB	267	GSTKQLVPFFLSTLNLECPSHNPNASVYIEVSCGEHGHTKRLDAIDNS	KDRVSSADYA	326
QY	329	---QNLGCAAEKKSSPEKN-----		345
DB	327	GLKARNDLVKVQNLKAILDKNDASSYSGRYEDNLTLNGLLNGWVNDI	VEKGVSSALV	386
QY	346	-----EVPAPCCPPEVPDPIESHTATSTLTQFCILFKRFTLSIL	RDTVLTH	393
DB	387	TTNERGDAMIDVEKSVNCTTALLTEETISPERYPTSQFHQFVWVLLKRT	LFLSFYRDWTLMY	446
QY	394	LRPMSHVVICVLITGLLYLHICDDASKVFENNTGCLFFFSMLFLMFAAL	MPVTLTFFPLEMAVF	453
DB	447	LRFLAHLVGLFGALYDYDIGNDGAKVLSNLGFLFFNMFLMYSMTITLS	FFLEPMVL	506
QY	454	MREHLNYSLSKAYYLAKTMADVPFQVCPVVCYSIVYMTGQPAETS	RRELLFSALAT	513
DB	507	LKENENRWYLSKSYLAISVADLPFK-----		LFSLSTCR 541
QY	514	ALV 516		
DB	542	SCI 544		

RESULT	15
Q8SZF5	
ID	Q8SZL
AC	Q8SZL
DT	01-JUN
DT	01-JUN

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobaray C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2193(2000).
CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AB003580; AAF51130.1; -
DR FlyBase; FBgn0031516; CG9663.
DR InterPro; IPR0031593; AAA_AtPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transport; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 812 AA; 90203 MW; CF3739101F12FB3E CRC64;

Query Match 34.8%; Score 1163.5; DB 5; Length 812;
Best Local Similarity 35.7%; Pred. No. 7.1e-78;
Matches 281; Conservative 119; Mismatches 235; Indels 153; Gaps 19;

QY 2 AKKALEVCGGLPGAVAMVT-----LEDGAEPPVLTHLK-----KVENHI--- 44
Db 29 SETHLPSSRGSGTGGALPRPTTPVPSKLLSSRRPPLHLTLNKAIRKSHDAPLAETVLTST 88
45 -----TEAQR-----FSLPKRS--AVDI 61
89 SESVGTPLGATVPSSPANQSORQASQSDSRNTSSTATSGGSIFFPHEQYTKTKINI 148
QY 62 EFVELSYVREGPCWKRKGKYLTKLCKSGKFCRRELIGIMPGSGAGKSTFMNLAGYRES 121
Db 149 GFENIRYTKFGVFORET--KDVLMGLTGYFKSGELSAVIGPSGAGKSTLLNLSGYTTY 206
QY 122 GKKGQILVNGRPRELRTFRKNSCYIMODMLPLHLTVLEAMVSNANLNTEN--PDVRND 179
Db 207 GFTGDFRNGNRDRDKAFKSNVAFTRQDTSLQAFLSVKEAMFAANLXIGTHTMSKRE 266
QY 180 LYTEILTLALGLMSCSHRTALLSGGQRKRLATALLVNNPPVMFTDEPTSGLDSDSCFOV 239
Db 267 RYKCLILEAIGYENRHTTGLSGGQKRLALALEVNNPPVLLIDPTTGLDSTSNQL 326
QY 240 VSLMSLAQGGRTIITCIHPSAKLFEMFDKLYILSOGQICPKGVVTVNLIPYKGLGLGHC 299
Db 327 INLLKKALEGRTVICTIHPSALTFAMFDHLYAIGCKCIYAGGAQNLLPLGLNALHC 386
QY 300 PYYHNPADF---IIEVASGEYGDNL-----PMLFRAVQNG-----LCAMA 336
Db 387 PESYNPADYCEFTVMEIATHDVTADNQLKLVALMDNGRNEYRQSKTARVAOLAAM- 445
QY 337 EKK-----SSP-EKNEVPA-----PCPP----- 353
Db 446 -KKVDQLMAAGLITPVTPAVMSTSVPAHMQGNTKFLTPINELSSRWVDSOTAGIGNAG 504

QY 354 --CPP-----EVDPIESH-----TFATSTLTQFCILFKRTFLSILRDTVLTHLR 395
Db 505 SCKPKKKKKNSRPAIEIDP--SHLCKRQNIYATPFYRQLSILLVRFLLIWRDSSLTWTR 562
QY 396 FNSHVIVGVILGLLHIGDDASKVFNNTGCLFSLFMLFMAALMPTVTFPLEMAVPMR 455
Db 563 FAHLITGLTGLTYFGIGNDAQTINIPRYLFYTIMFYMCAFGSILVKFPFPIVSR 622
QY 456 EHLNMYSLKAYYLAKTMADVFPQVVCVYVWVTGQPAETSRLFLFSALATATATL 515
Db 623 EHFNRWYSURAYIVALTADLPIQLICSALFIVPYLYMTQQPLELWRFQMFLLIVFTAL 682
QY 516 VAQSLGLLIGAASNSIQVATFVQVTAIPVLFSGFFVSEFKTIPYLOWSSYLSVYRGE 575
Db 683 VQSQIGLAVGAAL-SLKLSGILGPFPCFLOFSGFFLEKMDAPVFLRMWDFISFLKYSU 741
QY 576 EGVLITYGMERGDUITCLERCPFPREPQSIILRALDVEDAKLYMDFLVLGIFFLALRLAY 635
Db 742 EGATMAIFGYDREPLACNELYCHLRHPOFILKSLDMANGNYTLALIFLFGVLVFLRLAF 801
QY 636 LVLYRVK 643
Db 802 YIMSFLR 809

RESULT 14
QYVQNA
ID QYVQNA PRELIMINARY; PRT; 609 AA.
AC QYVQNA;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG9664 protein.
GN CG9664.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Adavani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobaray C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

RESULT 12
Q9VQJ9 PRELIMINARY; PRT: 620 AA.
AC Q9VQJ9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG3164 protein.
GN BEST:CK02656 OR CG3164.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K.E., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck G.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AE003590; AAF51548.1;
DR FlyBase: FBgn0025683; BEST:CK02656.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 620 AA; 70891 MW; 4EA5A336483E7BC5 CRC64;

Query Match 37.1%; Score 1241.5; DB 5; Length 620;
Best local Similarity 40.6%; Pred. No. 8.5e-84;
Matches 261; Conservative 103; Mismatches 184; Indels 95; Gaps 11;

QY 46 EAQFSLPKRSADVIEFVLSYSVREG-----PCWRKRGYTKLLKLSGKF 92
Db 31 QPKTLOHLPKRPAYDLAFHNLTRYKGNRSNHSSEGTOWTPPVWRHNGTIGCWQ-EY 89

```

[illegible]

```
Db 266 EITELIGKESVNTLTCNSGQKRLSIALELVNPNVMEFDEPTSGDSTSCFQLISL 325
QY 243 MKSLAQGRTICTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLIPYLKGLGHCHPTY 302
Db 326 LRSLARGRTIVCTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLIPYLKGLGHCHPTY 385
QY 303 HNPADFIIEVASGEYGLDNPMLFRAVQNGLC----- 333
Db 386 HNPADYVLEVASGEYGLDNPMLFRAVQNGLC----- 360
QY 334 --AMA-----EKKSPKNEVPAPCCP--PPEVDP----- 360
Db 446 ENAMAILTEDEKPPLEDROLESPTPVEDPADVPKPLETOQSONSDCSVVMNPTNAVDD 505
QY 361 -----IESHT-----PATSTLTQFCILF 378
Db 506 SCSFSSSGTQNAVGGSGGSGPSAVVGCMTSLDSDSHSVTLPNKGTGPTSGWTFWILL 565
QY 379 KRTFLSILRDVTLTHLRFSMHWVIGVLIGLGLYHLIGDADASKVFNTGCLFMSLFLMFAA 438
Db 566 KRSFTILRDKMLTHMRFLSHVIVGAIIGMIYDVGNEASKIMSAGCIPFVSLFTTFTA 625
QY 439 LMPVTTLTTPLENAVPMRHLNHWYSLKAYIYLAKTMAVDFVQVVCVVYVCSIVYVMTGQPA 498
Db 626 MNPVTTLTTPLENAVPMRHLNHWYSLKAYIYLAKTMAVDFVQVVCVVYVCSIVYVMTGQPA 685
QY 499 ETSRFLFSALATATATVAQSLGILLIGAASNSLOVATFVGPVTAIPVLLFSGFFVSFKTI 558
Db 686 ELERSVMTFLVCLNSLVAQSLGLLIGAGMN--IEFGVFLGPPVTIITILFSGFFVNFDTI 744
QY 559 PTLQWSSYLSVRYRGFEGVILTYIGMERGDLTCLERCPREPOSILRALDVEDAKLYM 618
Db 745 PCLQWVYVSVRYRGFEGVILTYIGMERGDLTCLERCPREPOSILRALDVEDAKLYM 804
QY 619 DELVL-GIFFALRLALVLYRYV 642
Db 805 DAVALIGIFF-ALRIAYFVLRWKL 828
RESULT 9
Q8R055 PRELIMINARY: PRT: 335 AA.
AC Q8R055
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Hypothetical 37.2 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
NCBI_TaxID=10090;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026477; AAH26477.1;
KW Hypothetical protein.
FT NON-TER
SQ SEQUENCE 335 AA; 37230 MW; C03AD5FB0D3D9EDD CRC64;
Query Match 40.8%; Score 1364; DB 11; Length 335;
Best Local Similarity 80.4%; Pred. No. 3.4e-93;
Matches 267; Conservative 13; Mismatches 34; Indels 18; Gaps 3;
QY 196 TTTALLSGQKRLAIALELVNPNVPMV---FFDEPT-----SGLDSASCFQVYSL 242
Db 2 TKTNOFRGQGRGAFPLVPSLSQKPPRLTHPLFLAISTSLVPGDLSGLDSASCFQVYSL 61
QY 243 MKSLAQGRTICTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLIPYLKGLGHCHPTY 302
Db 62 MKSLARGRTIVCTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLIPYLKGLGHCHPTY 121
```

```
QY 303 HNPADFIIEVASGEYGLDNPMLFRAVQNGLCAMAEKSSPEKNEVPAPCCPPEVDPIE 362
Db 122 HNPADFIIEVASGEYGLDNPMLFRAVQNGLCMAEKSSPGKNELPACHTCPCPELDPIE 181
QY 363 SHTFATSTLTQFCILFRKRTFLSILRDVTLTHLRFSMHWVIGVLIGLGLYHLIGDADASKYFN 422
Db 182 SHTFATSTLTQFCILFRKRTFLSILRDVTLTHLRFSMHWVIGVLIGLGLYHLIGDADASKYFN 241
QY 423 NTGCLFFSMLFLMFAALMPTVLTTPLENAVPMRHLNHWYSLKAYIYLAKTMAVDFVQVVC 482
Db 242 NTGCLFFSMLFLMFAALMPTVLTTPLENAVPMRHLNHWYSLKAYIYLAKTMAVDFVQVVC 301
QY 483 PVVYCSIVYVWMTG-----QPAETSRFLFSAL 509
Db 302 PVVYCSIVYVWMTG-----QPAETSRFLFSAL 333
RESULT 10
Q9VL61 PRELIMINARY: PRT: 689 AA.
AC Q9VL61
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE CG5853 protein.
GN CG5853
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jafarli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Liang D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.N., Nelson D.L.,
RA Palazzolo M., Pittman K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
```

[illegible]


```

RESULT 5
Q9NT30          PRELIMINARY;      PRT;      416 AA.
ID Q9NT30;
AC Q9NT30;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Hypothetical 46.6 kDa protein (Fragment).
GN DKFPZ434P1420.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: ALJ37563; CAB70814.1;
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 416 AA; 66574 MW; 6FA5984FEA38A1E6 CRC64;
Query Match 64.68; Score 2161; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 3,7e-152;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 231 LDSASCFQVVSMLKSLAOGGRTICTTHOPSAKLFEMFDKLYILSOGCIEFKGVVNTLIP 290
Db 1 LDSASCFQVVSMLKSLAOGGRTICTTHOPSAKLFEMFDKLYILSOGCIEFKGVVNTLIP 60
Qy 291 YLKGGLGHCTPHNPADFIIEVASGEYGDNLPMFLFRAVQNGLCMAEKKSSPEKNEVPAP 350
Db 61 YLKGGLGHCTPHNPADFIIEVASGEYGDNLPMFLFRAVQNGLCMAEKKSSPEKNEVPAP 120
Qy 351 CPPCPPEVDPIESHTFTATSTLTQFCILFKRTFLSILRDTVTLHFRMSHVVGVLIGLLY 410
Db 121 CPPCPPEVDPIESHTFTATSTLTQFCILFKRTFLSILRDTVTLHFRMSHVVGVLIGLLY 180
Qy 411 LHIGDDASKVFNNTGCLPFESMLFLMFAALMPTVTLTFLEMAVFMREHLNYYWSLKAYILA 470
Db 181 LHIGDDASKVFNNTGCLPFESMLFLMFAALMPTVTLTFLEMAVFMREHLNYYWSLKAYILA 240
Qy 471 KTMADVFPQVVCPPVYCSIVVWMTGQPAETSRFLFSALATATATVAQSLGLLIGCAASNS 530
Db 241 KTMADVFPQVVCPPVYCSIVVWMTGQPAETSRFLFSALATATATVAQSLGLLIGCAASNS 300
Qy 531 LQVATFVGPTAIPVLVLSGFFVSFKTIPTIYLOWSSYLSVRYGFEGLTIYGMERDGL 590
Db 301 LQVATFVGPTAIPVLVLSGFFVSFKTIPTIYLOWSSYLSVRYGFEGLTIYGMERDGL 360
Qy 591 TCLERCPFERPQSIIRALDVEDAKLYMDFVLGIFFLALLLALVLYRYRVKSER 646
Db 361 TCLERCPFERPQSIIRALDVEDAKLYMDFVLGIFFLALLLALVLYRYRVKSER 416

RESULT 6
Q8RIU5          PRELIMINARY;      PRT;      423 AA.
ID Q8RIU5;
AC Q8RIU5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 47.5 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

```

DR	EMBL; BC023077; AAH23077.1; -
KW	Hypothetical protein.
FT	NON_TER 1
SQ	SEQUENCE 423 AA; 47464 MW; 9F3F2FB5B5117FDD CRC64;
 Query Match 61.7%; Score 2064; DB 11; Length 423; Best Local Similarity 94.7%; Pred. No. 5.8e-145; Matches 396; Conservative 11; Mismatches 11; Indels 0; Gaps 0;	
QY	229 SGLDSASCFOVVS LKMSLAOAGRTTICTIHOPSAKLPMFDKLYLTLSOGCIFIKGVVTNL 288 6 SGLDSASCFOVVS LKMSLAHGGRIVCTIHOPSAKLPMFDKLYLTLSOGCIFIKGVVTNL 65
DB	289 IPYLKGLGLCHCPTYHNPADETFIEVASGEYGDLNPLMFLPRAVQNGLCMAEKSSPEKNVP 348
QY	66 IPYLKGLGLCHCPTYHNPADETFIEVASGEYGDLNPLMFLPRAVQNGLCTMAEKKSSPKNELP 125
DB	349 APCPPCPPEVDPIESHTFATSTLTQFCILFKRTELSILRDVTVLTHLRPMSHVIGLVTL 408
QY	126 AHCPTCPPELDPIESHTFATSTLTQFCILFRTELSILRDVTVLTHLRPMSHVIGLVTL 185
DB	409 LYHLHGDDASKVFNTGCLFSMLFLMPAALMPVLTLPLEMAVMFHRLANYSLKAY 468
QY	186 LYHLHGDDASKVFNTGCLFSMLFLMPAALMPVLTLPLEMAVMFHRLANYWTLKAY 245
DB	469 LAKTMADVPFOVCPPVVCYSIVYWMTPGPAETSRFLFSALATAVALQAOSLLGLIGAAS 528
QY	246 LAKTMADVPFOVCPPVVCYSIVYWMTPGPAETSRFLFSALATAVALQAOSLLGLIGAAS 305
DB	529 NSLOAVTVFGVTAIPVLLFSGFFVSFKTIPTYLOWSSYLSVRYGFEGVILTGYMERG 588
QY	306 TSLQAVTVFGVTAIPVLLFSGFFVSFKTIPTYLOWSSYLSVRYGFEGVILTGYMERG 365
DB	589 DLTCLERCPRPQSITILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLYRVKSR 646
QY	366 HLTCLDQCPRPDPOIILRELVDVEAKLYMDFLVLGIFFLALRLAYLVLYRVKSR 423
 RESULT 7	
Q9VQV4	PRELIMINARY; PRT; 832 AA.
ID	Q9VQV4
AC	Q9VQV4;
RX	STRAIN=BERKELEY;
RC	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA	George R.A., Lewis S.E., Richards S.J., Ashburner M., Henderson S.N.,
RA	Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Becson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Crowley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
OX	NCB1_taxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RE	STRAIN=BERKELEY;
RC	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA	George R.A., Lewis S.E., Richards S.J., Ashburner M., Henderson S.N.,
RA	Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Becson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Crowley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,


```
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
DR ProSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 646 AA; 72097 MW; E8344C4301ED6FB CRC64;

Query Match 95.7%; Score 3202; DB 11; Length 646;
Best Local Similarity 95.5%; Pred. No. 4.2e-229;
Matches 617; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAYD 60
DB 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAYD 60
QY 61 IEFVELSVREGPCWRKRGVYKTLKLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRE 120
DB 61 IEFVELSVREGPCWRKRGVYKTLKLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRE 120
QY 121 SGMKGOILVNGRPRELRTFRKMSCYIMODDMLLPHLTVLEAMVYSANLNLTENPDVKNDL 180
DB 121 SGMKGOILVNGRPRELRTFRKMSCYIMODDMLLPHLTVLEAMVYSANLNLTENPDVKNDL 180
QY 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPPVNFDEPTSGLDASCFQV 240
DB 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPPVNFDEPTSGLDASCFQV 240
QY 241 SLMKSIAOGGRTIICTIHQPSAKLFEMEDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCP 300
DB 241 SLMKSIAOGGRTIICTIHQPSAKLFEMEDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCP 300
QY 301 TYHNPADRIIEVASGEYGDLPMPFRAVONGLCMAEKKSPKNEVPAPCCPEVD 360
DB 301 TYHNPADRIIEVASGEYGDLPMPFRAVONGLCMAEKKSPKNEVPAPCCPEVD 360
QY 361 IESHTFATSTQFCILKPRFLSLTRDTVTLHFRFMSHVIGVLIHLHIGDDASKV 420
DB 361 IESHTFATSTQFCILKPRFLSLTRDTVTLHFRFMSHVIGVLIHLHIGDDASKV 420
QY 421 FNNTGCLFFSMLFLMFAALMPTVTLFPLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQV 480
DB 421 FNNTGCLFFSMLFLMFAALMPTVTLFPLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQV 480
QY 481 VCPVYVCIVVMTGQPAETSRFLFSALATATATVAOSGLLIGAAASNSLQVATFVGPV 540
DB 481 VCPVYVCIVVMTGQPAETSRFLFSALATATATVAOSGLLIGAAASNSLQVATFVGPV 540
QY 541 TAIPVLLSGFVFSEKTIPTIYLOWSSYLSYVRYGEGVILTIYGMERGLTCLERCPFR 600
DB 541 TAIPVLLSGFVFSEKTIPTIYLOWSSYLSYVRYGEGVILTIYGMERGLTCLERCPFR 600
DB 601 EPQSTLRALDVEDAKLYMDFLVGLGIFFLALRLAYLVLYRVYKSER 646
DB 601 DPQIILRELDEVEAKLYMDFLVGLGIFFLALRLAYLVLYRVYKSER 646

RESULT 2
Q91WA9 PRELIMINARY; PRT; 627 AA.
AC Q91WA9;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE Similar to ATP-binding cassette, sub-family G (White), member 4.
GN 6430517004RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE, AND RETINA;
RA Strausberg R.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC016200; AAI16200.1; -.
DR MGI: MGI:1924137; 6430517004RIK.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR ProSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 627 AA; 70371 MW; 28684468CB125D64 CRC64;

Query Match 92.8%; Score 3106; DB 11; Length 627;
Best Local Similarity 95.4%; Pred. No. 5.3e-22;
Matches 598; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 20 MAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAYDIEFVELSVREGPCWRKR 79
DB 20 MAVTLEDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAYDIEFVELSVREGPCWRKR 79
QY 80 GYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRESGMMKGQILVNGRPRELRTF 139
DB 80 GYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRESGMMKGQILVNGRPRELRTF 139
QY 140 RMKSCYIMODDMLLPHLTVLEAMVYSANLNLTENPDVKNDLVTETLTALGLMSCSHTRTA 199
DB 140 RMKSCYIMODDMLLPHLTVLEAMVYSANLNLTENPDVKNDLVTETLTALGLMSCSHTRTA 199
QY 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPPVNFDEPTSGLDASCFQV 259
DB 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPPVNFDEPTSGLDASCFQV 259
QY 260 PSAKLFEMEDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCPHYHNPADRIIEVASGEYGD 319
DB 260 PSAKLFEMEDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCPHYHNPADRIIEVASGEYGD 319
QY 320 LNPMLFRAVONGLCMAEKKSPKNEVPAPCCPEVDPTESHTATSTLTQFCILFK 379
DB 320 LNPMLFRAVONGLCMAEKKSPKNEVPAPCCPEVDPTESHTATSTLTQFCILFK 379
QY 380 RTFSLIRDTVTLHFRFMSHVIGVLIHLHIGDDASKVFNNTGCLFFSMLFLMFAAL 439
DB 380 RTFSLIRDTVTLHFRFMSHVIGVLIHLHIGDDASKVFNNTGCLFFSMLFLMFAAL 439
QY 440 MPTVTLFPLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQVVPVYVCIVVMTGQPAE 499
DB 440 MPTVTLFPLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQVVPVYVCIVVMTGQPAE 499
QY 500 TSRELLFSALATATATVAOSGLLIGAAASNSLQVATFVGPVTAIPVLLSGFVFSEKTIPT 559
DB 500 TSRELLFSALATATATVAOSGLLIGAAASNSLQVATFVGPVTAIPVLLSGFVFSEKTIPT 559
QY 560 TYLOWSSYLSYVRYGEGVILTIYGMERGLTCLERCPFRPQSTLRALDVEDAKLYMD 619
DB 560 TYLOWSSYLSYVRYGEGVILTIYGMERGLTCLERCPFRPQSTLRALDVEDAKLYMD 619
QY 620 FLVGLGIFFLALRLAYLVLYRVYKSER 646
DB 620 FLVGLGIFFLALRLAYLVLYRVYKSER 646
DB 601 FLVGLGIFFLALRLAYLVLYRVYKSER 627.

RESULT 3
Q9EPG9 PRELIMINARY; PRT; 666 AA.
AC Q9EPG9;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DE Similar to ATP-binding cassette, sub-family G (White), member 4.
GN 6430517004RIK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:16:04 ; Search time 87 Seconds
(without alignments)
1529.959 Million cell updates/sec

Title: US-10-072-621-9
Perfect score: 3347
Sequence: 1 MAEKALEAVGGGLPGAVAM.....FLALRLLAYLVLRVRSER 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

rchd: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

(SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rviris.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3202	95.7	646	11 Q8VBS9	Q8vbs9 mus musculus
2	3106	92.8	627	11 Q91WA9	Q91wa9 mus musculus
3	2360	70.5	666	11 Q9EPG9	Q9epg9 rattus norv
4	2177	65.0	785	4 Q96L76	Q96l76 homo sapien
5	2161	64.6	416	4 Q9NT30	Q9nt30 homo sapien
6	2064	61.7	423	11 Q8RIU5	Q8riu5 mus musculus
7	1723	51.5	832	5 Q9VQY4	Q9vqy4 drosophila
8	1723	51.5	832	5 Q9UAF0	Q9uaf0 drosophila
9	1364	40.8	335	11 Q8R0SS	Q8r0ss mus musculus
10	1322	39.5	689	5 Q9VL61	Q9vl61 drosophila
11	1319	39.4	689	5 Q96OD7	Q96od7 drosophila
12	1241.5	37.1	620	5 Q9VPJ9	Q9vpj9 drosophila
13	1163.5	34.8	812	5 Q9VQNS	Q9vqns drosophila
14	1086.5	32.5	609	5 Q9VQNM	Q9vqnm drosophila
15	1079.5	32.3	576	5 Q8SZF5	Q8szf5 drosophila
16	1070	32.0	615	5 Q8T998	Q8t998 drosophila

ALIGNMENTS

RESULT 1

Q8VBS9	PRELIMINARY;	PRT;	646 AA.
ID Q8VBS9			
AC Q8VBS9;			
DT 01-MAR-2002 (TRENBLrel. 20, Created)			
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)			
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)			
DE Putative white family ABC-transporter (ATP-binding cassette transporter sub-family G member 4) (ATP-binding cassette transporter ABCG4).			
DE ABCG4.			
GN ABCG4.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=RETINA;			
RA Oldfield S., Lowry C.A., Lightman S.L.;			
RT "Cloning and expression of a mammalian white family ABC-transporter, ABCG4."			
RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE FROM N.A.			
RA Annino T., Tammur J., Hutchinson A., Rzhetsky A., Dean M., Allikmets R.;			
RT "Human and mouse orthologs of a new ATP-binding cassette gene, ABCG4."			
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
RN [3]			
RP SEQUENCE FROM N.A.			
RC STRAIN-BALB/C; TISSUE=BRIN;			
RA Yoshikawa M., Yabuuchi H., Ikegami Y., Ishikawa T.;			
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AJ426047; CAD19779.2; -			
DR EMBL; AF411084; CAD137369.1; -			
DR EMBL; AY040865; AAK91781.1; -			
DR InterPro; IPR003593; AAA_ATPase.			
DR InterPro; IPR003439; ABC_transportr.			

17	1067	31.9	615	5	Q9VQF1	Q9vqf1 drosophila
18	1060	31.7	766	5	Q8T9E6	Q8t9e6 drosophila
19	1051	31.4	788	5	Q9VQ41	Q9vq41 drosophila
20	1016.5	30.4	643	5	Q9VPJ7	Q9vpj7 drosophila
21	950.5	28.4	232	4	Q43576	Q43576 homo sapien
22	921.5	27.5	623	5	Q8SXX6	Q8sxx6 drosophila
23	888	26.5	597	5	Q9VTL3	Q9vtl3 drosophila
24	873.5	26.1	687	5	Q94960	Q94960 drosophila
25	845.5	25.3	687	5	Q9NH94	Q9nh94 bombyx mori
26	820	24.5	679	5	Q9BH97	Q9bh97 ceratitis c
27	815.5	24.4	670	5	Q77423	Q77423 bactrocera
28	783	23.4	669	5	Q8WRE2	Q8wrf2 tribolium c
29	782.5	23.4	801	5	Q8T691	Q8t691 dictyosteli
30	777	23.2	669	5	Q8WRR1	Q8wrr1 tribolium c
31	770	23.0	672	10	Q9LI82	Q9li82 arabidopsis
32	767.5	22.9	648	10	Q9C6W5	Q9c6w5 arabidopsis
33	764.5	22.8	692	5	P91892	P91892 aedes aegypt
34	760.5	22.7	646	10	Q9C6R7	Q9c6r7 arabidopsis
35	719	21.5	657	11	Q9R004	Q9r004 mus musculu
36	717	21.4	655	4	Q96LD6	Q96ld6 homo sapien
37	716.5	21.4	658	5	O16574	O16574 caenorhabdi
38	712	21.3	655	4	Q96TA8	Q96ta8 homo sapien
39	705.5	21.1	662	10	Q949Y4	Q949y4 arabidopsis
40	699.5	20.9	609	10	Q9C8W6	Q9c8w6 arabidopsis
41	697	20.8	737	10	Q9FT51	Q9ft51 arabidopsis
42	691.5	20.7	751	10	Q93Y54	Q93y54 arabidopsis
43	689	20.6	668	10	Q9ARU4	Q9aru4 oryza sativ
44	686	20.5	695	5	Q19585	Q19585 caenorhabdi
45	685.5	20.5	1476	5	Q965D3	Q965d3 dictyosteli

[illegible]

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;

```

```

:
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/772,270A
:
: FILING DATE: December 23, 1996
:
: CLASSIFICATION: 424
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Gravelle, Micheline
:
: REGISTRATION NUMBER: 40,261
:
: REFERENCE/DOCKET NUMBER: 6580-81
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (416) 364-7311
:
: TELEFAX: (416) 361-1398
:
: INFORMATION FOR SEQ ID NO: 12:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 711 amino acids
:
: TYPE: amino acid
:
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: ORIGINAL SOURCE:
:
: ORGANISM: Actinobacillus pleuropneumoniae
:
: US-08-772-270A-12
:
:
:
: Query Match 6.9%; Score 230.5; DB 3; Length 711;
:
: Best Local Similarity 26.3%; Pred. No. 1.8e-14;
:
: Matches 82; Conservative 54; Mismatches 103; Indels 73; Gaps 10;
:
:
: QY 18 VAMAVTLEDGAEPVLTTHLKKVENHITEAQR-----FSLPKRSADVIEFV 64
:
: Db 429 VRLAQLWQDFQVGISITRLGDVLSPTENYQCKLSLPEIFGDIAPKHIRFYKPDAPII 488
:
: QY 65 --ELSVYRGEPCWRKRGYKTLKCLSGFCRRELIGIMGPSGAGKSTFNWILAGYRESG 122
:
: Db 489 LDVNLVSQK-----EVIGIVGRSGSGKSTLTLLKLRQF-YIP 525
:
: QY 123 MKGQILVNGRPREL---RTFRKMSCYIMQDMLL-----PHLT---VLEAMM 163
:
: Db 526 ENGQVLIDGHDALADPNWLRIGVGVLDQNVLLNSIRDNIALTDPMSMERVIYAACL 585
:
: QY 164 VSANLNLNTENPDVKNDLVTEILTALGLMSCSHTRTALLSGQKRRLAIALELVNPPVMF 223
:
: Db 586 AGAHDPISELREGYNTIVGEL-----GAGLSGQGRQRTAARALVNNPRILI 632
:
: QY 224 FDEPTSGLSASCFQVVYSLMKSLAQGGRTTICTIHOPSAKLFEMFDKLYLTLSOGOCIFKG 283
:
: Db 633 FDEATSALDYSEHITMQNMOKICH-GRVILIIAHLST--VKNADRIIVMEKHIVEQ 689
:
:
:
: 284 VTNLPIYLKGL 295
:
: Db 690 KHNLLNENGL 701

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402.804
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/114,420
; FILING DATE: 30-AUG-1993
; APPLICATION NUMBER: US 08/112,387
; FILING DATE: 27-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,928
; FILING DATE: 08-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/612,330
; FILING DATE: 13-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gray III, William O.
; REGISTRATION NUMBER: 30,944
; REFERENCE/DOCKET NUMBER: P/1261-13
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-402-804-6

Query Match 6.9%; Score 231.5; DB 2; Length 242;
Best Local Similarity 28.9%; Pred. No. 2.5e-15;
Matches 61; Conservative 46; Mismatches 95; Indels 9; Gaps 135

QY 80 GYKTLKCLSGKRCRRELIGIMGPSGAGKSTFMNLAGYRESGMKGQILVN---GRPRE 135
Db 12 GTHVHLKIFNLVSVEGEKLYLIGPSGSKETTIRCNGLLEVS-SEGVVVNLYLNHNK 70
QY 136 LRTFRKSCYTMQDDMLLPHLTVLEAMVSAANLNLTENPDVKNLDLVTETLTALGLMSCSH 195
Db 71 IEICRKYCAMVFOHFNLYPHMTVLQNLTPAPMKLQKKSKKEAETAFKYLKVVGLLDKAN 130
QY 196 RTTALSGGQRKRAIALELVNPNPVMFDEPTSGLDSASCFOVVSMLKSLA-QGRTII 254
Db 131 VVPATLSGGQORVAIRASLCTKKPYILFDEPTSAIDPETIQEVLDMVKMEISHQSNNTMV 190
QY 255 CTIHOPS-AKLFMFDKLYILSGOCIFKGV 284
Db 191 VVTHEMGFAK--EVADRIIFMEDGAIVEENI 219

RESULT 12
US-08-772-270A-12
; Sequence 12, Application US/08772270A
; Patent No. 601984
; GENERAL INFORMATION:
; APPLICANT: MacInnes, Janet
; APPLICANT: Ricciatti, Paul
; APPLICANT: Mallard, Bonnie
; APPLICANT: Rosendal, Soren
; TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR
; TITLE OF INVENTION: PRODUCING SAME, AND THEIR USE AS VACCINES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bereskin & Parr
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2

```


Matches 60; Conservative 37; Mismatches 81; Indels 3; Gaps 2;
QY 80 GYKTLKCLSGKFCRRRELIGTMGPGAGKSTFNNILAGYRESGMKGQILVNGRPRELRTF 139
DB 12 GKOILSDFSLSIEPKQLALVGPSSGSKTLLRLWLAGL-ETIDSGQIFYNQPLELDEL 70
QY 140 RMSC--YIMODMLLPHLTVLEAMVSNLNTENPDVKNLVTILITLALGMLSCSHTR 197
DB 71 QKRNLGPFVDFQLFPHLSVLENTLSVPKTMGKQEEAEKASGLLEQLGLGHAESY 130
QY 198 TALLSGGORKRLALALELVNPNPFDEPTSGDLSASCFOVSLMKSLAOGGRTICTI 257
DB 131 PFSLSGGOKORVALARAMIDPEITIGYDEPTSLDPELRLEVERKLILQNRRELGMTQIVWT 190
QY 258 H 258
DB 191 H 191

RESULT 7
US-09-134-001C-3832
Sequence 3832, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3832
LENGTH: 242
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3832

Query Match 7.1%; Score 238.5; DB 4; Length 242;
Best Local Similarity 30.3%; Pred. No. 4.8e-16;
Matches 67; Conservative 45; Mismatches 94; Indels 15; Gaps 5;
QY 77 RRGYKTLKCLSGKFCRRRELIGTMGPGAGKSTF---MNLAGYRESGMKGQILVNGRP 133
DB 11 KKGANEVLRLDINTVKEGVVAIGSGSGKSTLLRCMNL---DVPSGKGVIFEDNE 66
134 R-----ELRTRKMSCYIMODMLLPHLTVLEAMVSNLNTENPDVKNLVTILITLAL 188
67 LTQNVHLDNLQRKMGVWFQNFNLPFKKVIENVMALPLLHLKDSKOLKEKALYILEKV 126
QY 189 GLMCSHTRTALLSGGORKRLALALELVNPNPFDEPTSGDLSASCFOVSLMKSLAQ 248
DB 127 GLKDKADSYNQLSGGOKQORVALARALAMEPDVMLDFDEPTSLDPELVVGVVGLKVMRQLAN 186
QY 249 GGRTICTIHOPS-AKLFEFMDKLYLSOGOCIFKGVVTNL 288
DB 187 ESMTMVIIVTHEMNFKA--EISDKVFMVADGVVVSGETPQNI 225

RESULT 8
US-09-134-001C-3215
Sequence 3215, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3215
LENGTH: 225
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3215

Query Match 7.1%; Score 236.5; DB 4; Length 225;
Best Local Similarity 28.9%; Pred. No. 6.9e-16;
Matches 65; Conservative 47; Mismatches 94; Indels 19; Gaps 6;

QY 61 IEFVELSYVREGPCWKRKGYKTLKCLSGKFCRRRELIGTMGPGAGKSTFNNILAGYRE 120
DB 7 LEFQVSKSFRDGNQMIQAVKPTDLKIQ-----ELIAIVGSGSGKSTFLT-MAGALQ 60
QY 121 SGMKGOILVNGRPRELRTFKMS-----CYIMODMLLPHLTVLEAMVSNLNTEN 173
DB 61 RPTSGNIYNNKNLSILSEKOLSOIRINEIGFILQSTNLVPLFIEQPKL---LGKYYK 117
QY 174 PDVKNLVTILITLALGMLSCSHTRTALLSGGORKRLALALELVNPNPFDEPTSGDLS 233
DB 118 DTLSEDEVQKLVKQLNSLSDITNQLPNOISGGQKQORVALAKAIYTNPSIILADEPTASLDT 177
QY 234 ASCFOVVS-LMKSLAOGGRTICTIHOPS-AKLFEFMDKLYLSOG 277
DB 178 NNAMAVMKILOKTKERNKTCVIVTH--DERLTHFCDKTVFEMDG 220

RESULT 9
US-09-134-001C-3731
Sequence 3731, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3731
LENGTH: 248
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3731

Query Match 7.0%; Score 235.5; DB 4; Length 248;
Best Local Similarity 28.5%; Pred. No. 1e-15;
Matches 70; Conservative 47; Mismatches 110; Indels 19; Gaps 5;

QY 55 KRSADVIEFVELSYVREGPCWKRKGYKTLKCLSGKFCRRRELIGTMGPGAGKSTFNNI 114
DB 1 KRVSIMIELSHIHKSFNE-----TEVTKIGDLKINQGEVVTLLIGRSGSGKTKLLRM 51
QY 115 LAGYRESGMKGQILVNGRPRELR-----TPRKMSCYIMODMLLPHLTVLEAMVSNL 168
DB 52 INAL-EIPTGTVYVNGWYNTYNTKKSOIKVQSGMVWFQYVNFPHKSALENWEGUIT 110
QY 169 NLTENPDVKNLVTILITLALGMLSCSHTRTALLSGGORKRLALALELVNPNPFDEPT 228
DB 111 VKKNKATANEANMLAKVGLVHVQORPHALSGGQORVALARALAMNPKVMLFDEPT 170
QY 229 SGLDSASCFOVVSIMKSLAOGGRTICTIHOPS-AKLFEFMDKLYLSOGOCIFKGVVTN 287
DB 171 SALDPELVNDVLKVIKELADEGMTMVIIVTHEMREK--EVSNQIAFTHIEGVIAEQGTPED 228

```

Matches      73; Conservative   55; Mismatches    85; Indels     51; Gaps     11
QY       32 VLTT-H-LKKVNHTEAQRFSLPKRSADVDFIEPVLSYSVSREGPCWKRKRGYKTLLKCLSG 90
DB         : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB        7 ILTANOLSKV-----YGHKQRYALN---NISFSVEKG-----NISSFVKG-   36
QY       91 KFCRRRELIGMGPAGKAGSTFMNILAGYRESGMKGGLLYNGRP-----RFLTFTRKMSC 144
DB          : | |||||:::||: |: | | | | | | | | | | | | | | | | | | | | |
DB       37 -----EFVAVMGPSGGKTKTLNLVLSSI-DTISGGTVVEVGKEINKLSHKEVANFRKHQL 90
QY      145 -YIMODMLLPHTLVLEAMMVSNLANLTENPDVNKDLDVTETILTALGELMSCSTRHALLSG 203
DB          : ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB       91 GFIFQDYSVLPTLTUVKENIMPLSVQKFHYEIQ-N-YKEVAEALGIYNLGNNKPSEIG 149
QY      204 GORKRLAIALELVNPNPVPFFDEPTSGDLSDASCFOWVSLMKSL-AOGGRTIICITIHOPSA 262
DB          : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      150 GOOORTAAARAFVHKFTIIFADEPTCALDSKSQAQDLLHLREDMKNKFNSTIMMTVDPSA 209
QY      263 KLUFEMFDKLVIILSQGC----IPFG 283
DB           : :: | | | | | | | | | | | | | | | | | | | | | | | | |
DB      210 ASYA--ERVIMLKGDGHSEIVQG 231
                                           : :: | | | | | | | | | | | | | | | | | |

RESULT 6
US-08-858-207A-315
; Sequence 315, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 315:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-315
Query Match              7.1%; Score 238.5; DB 4; Length 203;
Best Local Similarity   33.1%; Pred. No. 3.7e-16;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:43:24 ; Search time 26 Seconds
(without alignments)
731.046 Million cell updates/sec

Title: US-10-072-621-9

Perfect score: 3347

Sequence: 1 MAEKALPAGVGLPGAVAM.....FLALRLLYLVRYVKSR 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PTCUS-COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710	21.2	655	US-09-245-808-1	Sequence 1, Appli
2	255.5	7.6	244	US-08-919-573-2	Sequence 2, Appli
3	255.5	7.6	244	US-08-919-573-4	Sequence 4, Appli
4	247.5	7.4	270	US-09-134-001C-5561	Sequence 5561, Ap
5	243.5	7.3	256	US-09-134-001C-4600	Sequence 4600, Ap
6	238.5	7.1	203	US-08-858-207A-315	Sequence 315, App
7	238.5	7.1	242	US-09-134-001C-3832	Sequence 3832, Ap
8	236.5	7.1	225	US-09-134-001C-3215	Sequence 3215, Ap
9	235.5	7.0	248	US-09-134-001C-3731	Sequence 3731, Ap
10	233.5	7.0	341	US-09-134-001C-4042	Sequence 4042, Ap
11	231.5	6.9	242	US-08-402-804-6	Sequence 6, Appli
12	230.5	6.9	711	US-08-772-270A-12	Sequence 12, Appl
13	223	6.7	215	US-09-305-984-18	Sequence 18, Appl
14	223	6.7	215	US-09-073-341A-18	Sequence 18, Appl
15	222.5	6.6	236	US-09-134-001C-4595	Sequence 4595, Ap
16	222.5	6.6	707	US-08-772-270A-4	Sequence 4, Appli
17	220	6.6	1307	US-08-395-246C-2	Sequence 2, Appli
18	218	6.5	215	US-09-305-984-64	Sequence 64, Appl
19	218	6.5	224	US-09-305-984-72	Sequence 72, Appl
20	218	6.5	224	US-09-305-984-74	Sequence 74, Appl
21	218	6.5	224	US-09-305-984-76	Sequence 76, Appl
22	215.5	6.4	254	US-09-134-001C-4440	Sequence 4440, Ap
23	215	6.4	1334	US-08-996-545-2	Sequence 2, Appli
24	215	6.4	1334	US-09-328-320-2	Sequence 2, Appli
25	214.5	6.4	246	US-09-627-376-15	Sequence 15, Appl
26	213.5	6.4	382	US-09-134-001C-3439	Sequence 3439, Ap
27	212.5	6.3	273	US-09-134-001C-3641	Sequence 3641, Ap

Sequence 332, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 75, Appl
Sequence 4, Appl
Sequence 12, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 19, Appl
Patent No. 5206352
Sequence 5550, Ap
Sequence 3539, Ap
Sequence 2, Appl
Sequence 2, Appl
Sequence 3369, Ap
Sequence 89, Appl

ALIGNMENTS

RESULT 1

US-09-245-808-1

; Sequence 1, Application US/09245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; FILE REFERENCE: encodes it

; CURRENT APPLICATION NUMBER: US/09/245,808

; EARLIER APPLICATION NUMBER: 1999-02-05

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Human MCF-7/AdrVp cells

US-09-245-808-1

Query Match 21.2%; Score 710; DB 4; Length 655;

Best Local Similarity 27.8%; Pred. No. 2e+63;

Matches 177; Conservative 140; Mismatches 247; Indels 72; Gaps 16;

Qy	61	IEFVELSVSV--REG--PCWRKRGYKTLKCLSGKFCRRELIGIMGPSGAGKSTFMNLA	116
Db	37	LSFNICYRVKLKSGGFLPC-RKPEKEILSNING-IMKPGNLALGPTGGKSLDLVLA	94
Qy	117	GYRE-SGMKQGIQVNGRPRELRTFRKMSYIMQDDMLPHLTVLEAMMVSAANLNL--TEN	173
Db	95	ARKDPSGLSGDLVINGAPRP-ANFKCNSGYVQDDVVMGTLTYRENQFSAAURLATTMT	153
Qy	174	PDVKNDLVTTEILTALGLMSSHTRTAL-----LSGGORKRLALALELVNPPVFFDEPT	228
Db	154	NHEKNEINRVIOELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFDEPT	213
Qy	229	SGLDSASCFQVSVLMKSLAOGRTICTIHOPSAKLFEMFDKLVLSOGOCIFKGVVTVNL	288
Db	214	TGLDSSANAVLLLLKRMKSKQGRITIFSIHOPRYSIFKLFDSLTLLASGRUMFHGPAQEA	273
Qy	289	IPYLKGLGLHCHPTNPAFFIEVASGEYGDLPNMLPRAVQNGLCAMAEKKSPKNEVP	348
Db	274	LGVFESAGYHCEAYNPNPDFLDIINDG-----STAVALNREEDFKATEII	319
Qy	349	APCPPCPEVDPI-----ESH-----TATSTLTQFCI	376
Db	320	EPSQDKPLIEKLAELIYVNSFSYKETAELHOLSGGKKKKTKITVFKETISYTTTSFCHLRW	379